

**NONPROVISIONAL**  
**APPLICATION FOR UNITED STATES LETTERS PATENT**  
**FOR**  
***GINKGO BILOBA* LEVOPIMARADIENE SYNTHASE**

**BY**

**SEIICHI P.T. MATSUDA**  
**AND**  
**HALA G. SCHEPMANN**

10041007.010702

EXPRESS MAIL LABEL # ET575269165US

# GINKGO BILOBA LEVOPIMARADIENE SYNTHASE

## FIELD OF THE INVENTION

[0001] The present invention is directed to the fields of molecular biology, molecular genetics, and organic chemistry. Specifically, the present invention is directed to the cloning and characterization of at least one *Ginkgo biloba* sequence for biosynthesis of ginkgolides. More specifically, the present invention is directed to the cloning, characterization and expression of *Ginkgo biloba* levopimaradiene synthase.

## BACKGROUND OF THE INVENTION

[0002] The gymnosperm *Ginkgo biloba*, of the Conopsida class, Ginkgoales order, and Ginkgoaceae family, originated in Eastern China approximately 150 million years ago and is the sole living representative of its order (Schwarz and Arigoni, 1999; Benson, L., 1957; Chaw, *et al.*, 2000; Bowe, *et al.*, 2000). This hardy tree, termed a "living fossil" by Charles Darwin, is well-known for its ability to withstand harsh climate conditions and resist insect infestation (Major, R. T., 1967). The apparent lack of change over millions of years is presumably due to its long time span between generations; reproduction begins after 20 years of age and continues to 1000 years of age.

[0003] *G. biloba* is renowned as a potent herbal therapeutic that aids in the revascularization of ischemic tissue through improved microcirculation. *G. biloba* leaf extracts have been used for centuries to treat cerebrovascular and cardiovascular diseases, dementia, tinnitus, arthritis, and vertigo (Itil, *et al.*, 1995; Briskin, D. P., 2000). These beneficial pharmacological effects have been attributed, in part, to the ginkgolides, a unique series of diterpene molecules which are highly specific platelet-activating factor (PAF) receptor antagonists (Hosford *et al.*, 1990). Generation of PAF occurs during anaphylaxis or shock and leads to bronchoconstriction, contraction of smooth muscle, and reduced coronary blood flow, which are often fatal. The isomer known as ginkgolide B demonstrates the highest activity of the diterpenes and antagonizes all known PAF-induced membrane events. Furthermore, the American Medical Association recently endorsed the Chinese herb as a viable alternative to traditional approaches in the treatment of Alzheimer's disease. Recent studies report that the extract delayed the progression of dementia in approximately one third of the patients studied (Le Bars *et al.*, 1997).

10041007.010702

[0004] Ginkgolides were first isolated from the roots of the *Ginkgo* tree by Furukawa (1932) and later characterized by K. Nakanishi (1967) and Sakabe (1967); the elucidated structures were named Ginkgolides A, B, C, J, and M. In 1967, K. Okabe also established the presence of the ginkgolides in the leaves of the *Ginkgo* tree. Ginkgolides are biosynthesized from geranylgeranyl diphosphate, the universal diterpene precursor. These molecules contain a caged trilactone structure and display a rare *tert*-butyl group. Analogs are distinguished by the number and location of hydroxyl group substituents. Recently, the ginkgolides and bilobalide (a pentanorditerpenoid by-product of ginkgolide biosynthesis) were determined to have significant antifeedant activities toward insect larvae (Schwarz, M., 1994; Matsumoto, *et al.*, 1987).

[0005] Geranylgeranyl diphosphate (GGDP) (Schwarz and Arigoni, 1999) employed in ginkgolide biosynthesis is derived from isopentenyl diphosphate formed via the deoxyxylulose pathway. The proposed biosynthesis of the ginkgolides is initiated by protonation of GGDP to give labdadienyl diphosphate. Ionization of the allylic diphosphate moiety followed by a 1,4-hydrogen shift, methyl migration, and deprotonation yields levopimaradiene (Schwarz and Arigoni, 1999). The proposed intramolecular hydrogen shift was also observed in the biosynthesis of *Abies grandis* abietadiene synthase (AgAS) (Ravn *et al.*, 1998; Ravn *et al.*, 2000). Oxidation of ring C produces abietatriene, which is then transported from the plastid to the cytoplasm. The aromatic hydrocarbon undergoes further transformation in the endoplasmic reticulum by cytochrome P450-dependent monooxygenases to produce the ginkgolides (Schwarz and Arigoni, 1999) (FIG. 1).

[0006] Metabolic regulation studies of diterpene production in *G. biloba* seedlings indicate that ginkgolides are produced in the roots and are subsequently translocated to the leaves. Furthermore, diterpene hydrocarbon precursors were found exclusively in the roots and included levopimaradiene, palustradiene, abietadiene, pimaradiene, and abietatriene. Addition of cytochrome P450-dependent oxygenase inhibitors to the roots of seedlings resulted in full inhibition of oxygenation reactions along the pathway to the diterpenes. Abietatriene, the sole diterpene hydrocarbon obtained, was identified as the immediate precursor to the ginkgolides (Cartayrade *et al.*, 1997; Neau *et al.*, 1997).

[0007] Presently, commercial development of the ginkgolides as therapeutic agents has been hampered. Because these diterpenoids contain up to 12 stereocenters, 4

contiguous quaternary carbons, and 3 oxacyclic rings fused to 2 spiro carbocyclic rings, they present a formidable synthetic challenge. In spite of the topological and stereochemical complexity inherent to the ginkgolides, total syntheses of these unusually challenging targets have been achieved. In 1988, the first synthesis of (±)-ginkgolide A (38 steps, < 1% overall yield) and (±)-ginkgolide B (35 steps, < 1% overall yield) were reported (Corey and Ghosh, 1988; Corey *et al.*, 1988). Furthermore, ginkgolide B was converted to ginkgolide A in 6 steps and approximately 50% yield. More recently, (±)-ginkgolide B was synthesized in 26 steps and 3% total yield (Crimmins *et al.*, 1999). Although strategically impressive, these demanding routes require multiple transformations resulting in low yields that ultimately preclude commercial-scale production of the ginkgolides.

[0008] Current commercial ginkgolide production relies exclusively on extraction from *Ginkgo* trees, which accumulate low levels of the compound. In addition, the demand for this medicinal plant has increased at a rate of 26% per annum with 2,000 tons harvested annually (Masood, E., 1997). *G. biloba* plantations serve as the major source of the herbal extract and provide an average 1 to 7 mg/g dry weight ginkgolide from young trees (Balz, *et al.*, 1999). In an effort to increase diterpenoid content, *G. biloba* seedlings, plants, and trees were treated with metabolic inhibitors that target key branchpoints in isoprenoid biosynthesis downstream of GGPP synthesis (Huh, *et al.*, 1993). Presumably, inhibiting GGPP depleting pathways would increase the available concentration of GGPP, the natural diterpene substrate. Variable results were obtained with cycloartenol synthase inhibitors, ancymidol and AMO-1618. In contrast, application of fluridone (an inhibitor of carotenoid biosynthesis that blocks phytoene desaturation) yielded up to 78% more ginkgolides.

[0009] Extraction of the ginkgolides from *G. biloba* is known. U.S. Pat. No. 5,399,348 refers to a method for preparation of *Ginkgo biloba* extract in which the alkylphenol compounds are separated not by using chlorinated aliphatic hydrocarbon, but through a process of precipitation, filtration and multi step liquid-liquid-extractions. U.S. Pat. No. 5,399,348; U.S. Pat. No. 5,322,688; U.S. Pat. No. 5,389,370; U.S. Pat. No. 5,389,370; U.S. Pat. No. 5,637,302; U.S. Pat. No. 5,512,286; U.S. Pat. No. 5,399,348; and U.S. Pat. No. 5,389,370 are all directed to various methods of preparing a desired *Ginkgo*

*biloba* extract. U.S. Pat. No. 5,241,084 and U.S. Pat. No. 5,599,950 are directed to methods to convert ginkgolide C to ginkgolide B.

[0010] Seeking an alternative, non-synthetic approach to ginkgolide production, a method to clone and functionally express genes involved in their biosynthesis was considered. In 1971, the isoprenoid nature of the ginkgolides was precariously, yet correctly, established using 2-<sup>14</sup>C MVA incorporation experiments conducted with *G. biloba* seedlings. Moreover, the researchers proposed that the unique *tert*-butyl group arose from *S*-adenosyl methionine (Nakanishi, *et al.*, 1971). However, a revised biogenetic scheme was put forth as a result of NMR product analyses of isotopically labeled precursors incubated with *G. biloba* embryos (Schwarz, *et al.*, 1999). During the course of these extensive studies, a dichotomy was observed concerning the biosynthesis of IPP by *G. biloba*. Specifically, formation of isopentenyl pyrophosphate (IPP), an isoprene unit possessing a diphosphate moiety, proceeds via the classical MVA pathway in the synthesis of sitosterol, but in the plastids, the deoxyxylulose-5-phosphate (DXP) pathway synthesizes GGPP. Presumably, segregation between the two pathways is due to compartmentalization of the plant cell. IPP responsible for sitosterol formation is restricted to the cytoplasm, and IPP incorporated into ginkgolides originates in the chloroplasts.

[0011] *G. biloba* cultures were first established in 1971; however, HPLC analysis failed to detect ginkgolides (Nakanishi, *et al.*, 1971). Two decades later, ginkgolides A and B were detected in undifferentiated cell cultures (<20 ng/g dry weight), albeit by a factor of 10<sup>6</sup> less than that obtained from leaves of mature trees (Carrier, *et al.*, 1991; Chauret, *et al.*, 1991). Increased ginkgolide content was demonstrated in primary callus and cell suspension cultures (~26% and 47% relative to leaves of mature trees, respectively) were unable to be maintained in secondary cultures (Huh, *et al.*, 1993). Currently, high yield production of the ginkgolides by *in vitro* cultures of undifferentiated cells has not been achieved (Balz *et al.*, 1999). Transgenic cells were obtained from putative transformed *G. biloba* embryos but ginkgolide concentration was <400 µg/g dry tissue culture (Laurain, *et al.*, 1997). Recently, Dupré *et al.* (2000) reported a reproducible transformation protocol of *G. biloba* by *Agrobacterium tumefaciens*; however, ginkgolide levels of the transformed cells have not been disclosed.

[0012] There are examples in the art in which heterologous diterpene synthases are introduced into and expressed in organisms such as *Escherichia coli*, particularly for the purpose of characterizing activity of a soluble form of the enzyme in the absence of any plastidial targeting sequence (Hill *et al.*, 1996; Peters *et al.*, 2000; Williams *et al.*, 2000). However, the novel levopimaradiene synthase of the present invention provides a solution to a need in the art for methods and compositions to quickly produce large amounts of substantially pure ginkgolides in a cost-effective manner, particularly in an organism capable of a high-yield ginkgolide-producing system.

## SUMMARY OF THE INVENTION

[0013] One embodiment of the present invention is a purified and isolated nucleic acid sequence encoding a levopimaradiene synthase.

[0014] Another embodiment of the present invention is a purified and isolated nucleic acid sequence comprising, SEQ.ID.NO:1, SEQ.ID.NO:32, SEQ.ID.NO:34, SEQ.ID.NO:36 or SEQ.ID.NO:38.

[0015] An additional embodiment is a purified and isolated nucleic acid comprising SEQ.ID.NO:34. Another embodiment is a purified and isolated nucleic acid comprising SEQ.ID.NO:36.

[0016] Another embodiment of the present invention is an expression vector comprising an isolated and purified nucleic acid sequence encoding a levopimaradiene synthase under control of a promoter operable in a host cell. In a specific embodiment, the promoter is an inducible promoter, and preferably GAL1. In another specific embodiment, the nucleic acid sequence comprises SEQ.ID.NO:1, SEQ.ID.NO:32, SEQ.ID.NO:34, SEQ.ID.NO:36 or SEQ.ID.NO:38.

[0017] In yet another specific embodiment of the present invention, the host cell is a prokaryote, and preferably *Escherichia coli*. In another specific embodiment, the host cell is a eukaryote, and in a preferred specific embodiment the eukaryote is a yeast.

[0018] Another embodiment of the present invention is an isolated polypeptide having an amino acid sequence of a levopimaradiene synthase.

[0019] In another embodiment of the present invention there is an isolated polypeptide comprising an amino acid sequence of SEQ.ID.NO:2, SEQ.ID.NO:33, SEQ.ID.NO:35, SEQ.ID.NO:37 or SEQ.ID.NO:39.

[0020] Another embodiment of the present invention is an isolated polypeptide comprising an amino acid sequence of SEQ.ID.NO:35. Further, the present invention embodies an isolated polypeptide comprising an amino acid sequence of SEQ.ID.NO:37.

[0021] Another embodiment of the present invention is an expression vector comprising an isolated polynucleotide sequence encoding a polypeptide having an amino acid sequence of a levopimaradiene synthase. In a specific embodiment, the vector further comprises a promoter operatively linked to the polynucleotide sequence. In a further specific embodiment, the promoter is an inducible promoter. In a preferred specific embodiment, the inducible promoter is GAL1.

[0022] In another embodiment there is an isolated polynucleotide sequence encoding a polypeptide having an amino acid sequence of SEQ.ID.NO:2, SEQ.ID.NO:33, SEQ.ID.NO:35, SEQ.ID.NO:37 or SEQ.ID.NO:39. In a specific embodiment, the vector further comprises a promoter operatively linked to the polynucleotide sequence. In a further specific embodiment, the promoter is an inducible promoter and preferably GAL1.

[0023] In another embodiment of the present invention, there is a unicellular organism comprising a purified and isolated nucleic acid sequence encoding a levopimaradiene synthase. In a specific embodiment, the nucleic acid sequence comprises SEQ.ID.NO:1, SEQ.ID.NO:32, SEQ.ID.NO:34, SEQ.ID.NO:36 or SEQ.ID.NO:38. In a further specific embodiment, the nucleic acid sequence comprises an expression vector. In yet a further specific embodiment, the expression vector comprises an inducible promoter. In a preferred specific embodiment, the inducible promoter is GAL1.

[0024] In another specific embodiment of the present invention, the nucleic acid sequence encoding the levopimaradiene synthase contains a deletion corresponding to an N-terminal sequence. In yet another specific embodiment, the organism is *Saccharomyces*, *Escherichia coli*, *Candida albicans* or *Kluyveromyces lactis*. In a preferred specific embodiment, the organism is *Escherichia coli*. In another preferred specific embodiment, the organism is *Saccharomyces cerevisiae*.

[0025] Another embodiment of the present invention is a yeast host cell comprising a vector, wherein the vector comprises a purified and isolated nucleic acid sequence comprising SEQ.ID.NO:1, SEQ.ID.NO:32, SEQ.ID.NO:34, SEQ.ID.NO:36, or SEQ.ID.NO:38 wherein said nucleic acid sequence is under control of a promoter operable

in the yeast host cell. In a further specific embodiment, the nucleic acid sequence comprises an expression vector.

[0026] Yet another embodiment of the present invention is a yeast host cell comprising a vector, wherein the vector comprises an isolated polynucleotide sequence encoding a polypeptide having an amino acid sequence of SEQ.ID.NO:2, SEQ.ID.NO:33, SEQ.ID.NO:35, SEQ.ID.NO:37 or SEQ.ID.NO:39, wherein expression of the polynucleotide is under control of a promoter operable in the yeast host cell. In a further specific embodiment, the vector is an expression vector.

[0027] In one embodiment of the present invention there is a plant host cell, wherein the cell comprises an isolated and purified nucleic acid sequence comprising SEQ.ID.NO:1, SEQ.ID.NO:32, SEQ.ID.NO:34, SEQ.ID.NO:36 or SEQ.ID.NO:38, under control of a promoter operable in the yeast host cell. In a specific embodiment, the promoter is an inducible promoter. In another specific embodiment, the plant is *Ginkgo biloba*.

[0028] Another embodiment of the present invention there is a unicellular organism comprising an isolated polynucleotide sequence encoding a polypeptide having an amino acid sequence of a levopimaradiene synthase. In a specific embodiment the amino acid sequence comprises SEQ.ID.NO:2, SEQ.ID.NO:33, SEQ.ID.NO:35, SEQ.ID.NO:37 or SEQ.ID.NO:39. In another specific embodiment, the polynucleotide sequence contains a deletion corresponding to an N-terminal sequence of the levopimaradiene synthase.

[0029] In a specific embodiment, the unicellular organism is *Saccharomyces*, *Escherichia coli*, *Candida albicans*, or *Kluyveromyces lactis*. In other specific embodiments, the unicellular organism is *Saccharomyces cerevisiae* or *Escherichia coli*.

[0030] In one embodiment of the present invention there is a method of producing ginkgolide in a cell, comprising the steps of obtaining a cell comprising an isolated and purified nucleic acid sequence encoding a levopimaradiene synthase; culturing said cell under conditions wherein the cell produces ginkgolide; and removing the ginkgolide from the culture of cells. In a specific embodiment, the nucleic acid sequence comprises SEQ.ID.NO:1, SEQ.ID.NO:32, SEQ.ID.NO:34, SEQ.ID.NO:36 or SEQ.ID.NO:38. In a specific embodiment, the cell is *Saccharomyces cerevisiae*. In another specific embodiment, the cell is *Escherichia coli*. In a further specific embodiment, the nucleic



acid sequence comprises an expression vector, wherein the expression vector includes an inducible promoter operatively linked to the levopimaradiene synthase coding region.

[0031] In another embodiment of the present invention there is a method of producing levopimaradiene in a cell, comprising the steps of obtaining a cell comprising an isolated and purified nucleic acid sequence encoding a levopimaradiene synthase; culturing the cell under conditions wherein the cell produces levopimaradiene; and removing the levopimaradiene from the culture of cells. In a specific embodiment, the nucleic acid sequence comprises SEQ.ID.NO:1, SEQ.ID.NO:32, SEQ.ID.NO:34, SEQ.ID.NO:36 or SEQ.ID.NO:38. In a further specific embodiment, the nucleic acid sequence comprises an expression vector, wherein the expression vector includes an inducible promoter operatively linked to the levopimaradiene synthase coding region.

[0032] In another embodiment of the present invention there is a method of producing a ginkgolide in a yeast cell, comprising the steps of obtaining a cell wherein an isolated and purified nucleic acid sequence of SEQ.ID.NO:1, SEQ.ID.NO:32, SEQ.ID.NO:34, SEQ.ID.NO:36 or SEQ.ID.NO:38 under control of a promoter operable in the yeast cell has been added to the yeast cell; culturing the cell under conditions wherein the yeast cell produces the ginkgolide; and removing the ginkgolide from the culture of yeast cells. In a specific embodiment, the nucleic acid sequence further comprises an inducible promoter.

[0033] In another embodiment of the present invention there is a method of producing levopimaradiene in a yeast cell, comprising the steps of obtaining a yeast cell wherein an isolated and purified nucleic acid sequence of SEQ.ID.NO:1, SEQ.ID.NO:32, SEQ.ID.NO:34, SEQ.ID.NO:36 or SEQ.ID.NO:38 under control of a promoter operable in the yeast cell has been added to the yeast cell; culturing the yeast cell under conditions wherein the yeast cell produces the levopimaradiene; and removing the levopimaradiene from the culture of yeast cells. In a further embodiment, the nucleic acid sequence and the promoter comprise an expression vector.

[0034] In another embodiment of the present invention there is a method of producing levopimaradiene in a yeast cell, comprising the steps of obtaining a yeast cell wherein an isolated polynucleotide sequence encoding a polypeptide having an amino acid sequence of a levopimaradiene synthase under control of a promoter operable in the yeast cell has been added to the yeast cell; culturing the yeast cell under conditions wherein the

yeast cell produces the levopimaradiene; and removing the levopimaradiene from the culture of yeast cells.

[0035] In a specific embodiment, the promoter is an inducible promoter. In another specific embodiment, the amino acid sequence comprises SEQ.ID.NO:2, SEQ.ID.NO:33, SEQ.ID.NO:35, SEQ.ID.NO:37 or SEQ.ID.NO:39.

[0036] Another embodiment of the present invention is a method of producing levopimaradiene in a cell, comprising the steps of obtaining a yeast cell, wherein an isolated and purified nucleic acid sequence of SEQ.ID.NO:1, SEQ.ID.NO:32, SEQ.ID.NO:34, SEQ.ID.NO:36 or SEQ.ID.NO:38 under control of a promoter operable in the yeast cell has been added to the yeast cell and the yeast cell further comprises an increase in the effective amount of geranylgeranyl diphosphate; growing a culture of the yeast cells; and removing the levopimaradiene from the culture of yeast cells.

[0037] In another embodiment of the present invention, there is a ginkgolide, wherein said ginkgolide is obtained from production in a unicellular organism comprising a purified and isolated nucleic acid sequence encoding a levopimaradiene synthase.

[0038] In another embodiment of the present invention, there is a ginkgolide, wherein said ginkgolide is obtained from production in a unicellular organism comprising a purified and isolated nucleic acid sequence of SEQ.ID.NO:1, SEQ.ID.NO:32, SEQ.ID.NO:34, SEQ.ID.NO:36 or SEQ.ID.NO:38.

[0039] Another embodiment of the present invention is a ginkgolide, wherein the ginkgolide is obtained from production in a unicellular organism comprising an expression vector having an isolated and purified nucleic acid sequence encoding a levopimaradiene synthase under control of a promoter operable in the organism.

[0040] Another embodiment of the present invention is a ginkgolide, wherein the ginkgolide is obtained from production in a unicellular organism, wherein the organism comprises an isolated polynucleotide sequence encoding a polypeptide having an amino acid sequence of SEQ.ID.NO:2, SEQ.ID.NO:33, SEQ.ID.NO:35, SEQ.ID.NO:37 or SEQ.ID.NO:39.

[0041] In another embodiment of the present invention, there is a ginkgolide, wherein said ginkgolide is obtained from the method of producing the ginkgolide in a cell comprising the steps of obtaining a culture of cells wherein at least one cell comprises a purified and isolated nucleic acid sequence encoding a levopimaradiene synthase;

10041007-010702

culturing the cell under conditions wherein the cell produces the ginkgolide; and removing the ginkgolide from the culture of cells. In a specific embodiment, the nucleic acid sequence comprises SEQ.ID.NO:1, SEQ.ID.NO:32, SEQ.ID.NO:34, SEQ.ID.NO:36 or SEQ.ID.NO:38.

[0042] In another embodiment of the present invention, there is a ginkgolide, wherein said ginkgolide is obtained from the method of producing the ginkgolide in a yeast cell, comprising the steps of obtaining a culture of yeast cells, wherein at least one yeast cell comprises a purified and isolated nucleic acid sequence of SEQ.ID.NO:1, SEQ.ID.NO:32, SEQ.ID.NO:34, SEQ.ID.NO:36 or SEQ.ID.NO:38; culturing the yeast cell under conditions wherein the yeast cell produces the ginkgolide; and removing the ginkgolide from the culture of yeast cells.

[0043] In another embodiment of the present invention, there is a ginkgolide, wherein said ginkgolide is obtained from production in a unicellular organism which includes an isolated polynucleotide sequence encoding a polypeptide having an amino acid sequence of a levopimaradiene synthase, wherein the polynucleotide sequence comprises a deletion corresponding to an N-terminal sequence; culturing the cell under conditions wherein the cell produces the ginkgolide; and removing the ginkgolide from the culture of cells. In a specific embodiment, the amino acid sequence comprises SEQ.ID.NO:33, SEQ.ID.NO:35, SEQ.ID.NO:37 or SEQ.ID.NO:39.

[0044] In an additional embodiment of the present invention, there is a nucleic acid sequence comprising SEQ.ID.NO:5, SEQ.ID.NO:6, SEQ.ID.NO:7, SEQ.ID.NO:8, SEQ.ID.NO:9, SEQ.ID.NO:10, SEQ.ID.NO:11, SEQ.ID.NO:12, SEQ.ID.NO:29, SEQ.ID.NO:30, SEQ.ID.NO:31 or SEQ.ID.NO:40.

[0045] In an additional embodiment of the present invention there is a transgenic plant, wherein the plant comprises a purified and isolated nucleic acid sequence encoding a levopimaradiene synthase under control of a promoter operable in the plant. In a specific embodiment, the plant is *Ginkgo biloba*. In another specific embodiment, the nucleic acid sequence comprises SEQ.ID.NO:1, SEQ.ID.NO:32, SEQ.ID.NO:34, SEQ.ID.NO:36 or SEQ.ID.NO:38.

[0046] In another specific embodiment, there is a seed of the transgenic plant. In a preferred embodiment, the seed is *Ginkgo biloba*.

[0047] Other and further objects, features, and advantages are apparent and eventually more readily understood by reading the following specification and the accompanying drawings forming a part thereof, or any examples of the presently preferred embodiments of the invention given for the purpose of the disclosure.

## **BRIEF DESCRIPTION OF THE DRAWINGS**

[0048] The following drawings form part of the present specification and are included to further demonstrate certain aspects of the present invention. The invention may be better understood by reference to one or more of these drawings in combination with the detailed description of specific embodiments presented herein:

[0049] FIG. 1 depicts the biosynthesis of ginkgolide A from geranylgeranyl diphosphate (GGDP).

[0050] FIG. 2 illustrates the structure of an isoprene unit.

[0051] FIG. 3 illustrates the parent ginkgolide chemical structure.

[0052] FIG. 4 illustrates amino acid sequence alignment of plant sesquiterpene and diterpene synthases.

## **DESCRIPTION OF THE INVENTION**

[0053] It will be readily apparent to one skilled in the art that various embodiments and modifications may be made in the invention disclosed herein without departing from the scope and spirit of the invention.

[0054] As used in the specification, "a" or "an" may mean one or more. As used in the claim(s), when used in conjunction with the word "comprising", the words "a" or "an" may mean one or more than one. As used herein "another" may mean at least a second or more.

[0055] The technology of the present invention is related to the invention described in the U.S. Patent Application entitled, "Diterpene-Producing Unicellular Organism" filed on the same day and incorporated by reference herein.

### **I. Definitions**

[0056] The term "diterpene" as used herein is defined as a terpene compound comprised of four isoprene units to yield a 20-carbon hydrocarbon structure. The 20

carbon acyclic structure is called geranylgeranyl pyrophosphate (GGPP) or equally correct, geranylgeranyl diphosphate (GGDP). A skilled artisan is aware that diterpenes result from metabolism of GGPP and, thus may, after metabolism, yield a structure possessing one or more rings, one or more double bonds or one or more hydroxyl group. Non-limiting examples of diterpenes are levopimaradiene copalol, abietadiene and abietatriene.

[0057] The term “GGDP” as used herein is defined as geranylgeranyl diphosphate. The term may be used interchangeably with geranylgeranyl pyrophosphate (GGPP).

[0058] The term “GGPP” as used herein is defined as geranylgeranyl pyrophosphate. The term may be used interchangeably with geranylgeranyl diphosphate (GGDP).

[0059] The term “diterpenoid” as used herein is defined as a metabolite of a diterpene. One skilled in the art recognizes that a diterpene is often further transformed and, thus, may possess in an intermediate or final structure, more or less than the starting 20-carbons, one or more functional groups such as, for example, an ether, a carbonyl, an hydroxyl group or an aromatic ring.

[0060] The term “ginkgolide” as used herein is defined as a diterpenoid from the *Ginkgo biloba* plant. A skilled artisan is aware that there are at least the following naturally occurring ginkgolides: Ginkgolide A, Ginkgolide B, Ginkgolide C, Ginkgolide M, and Ginkgolide J. A skilled artisan is also aware that there are additionally many derivatives thereof, such as, for example, a ketone (*i.e.*, an acetate) at least one of any of the R groups in FIG. 3. A skilled artisan is aware that functional groups are often altered on a structure to effect characteristics such as, for example, solubility, and is very important in developing, for example, efficacious pharmaceuticals and medicaments.

[0061] The term “gymnosperm” as used herein is defined as a plant whose seeds are not enclosed within an ovary. Gymnosperms are contained in four phyla: Cycadophyta, Ginkgophyta, Pinophyta, and Gnetophyta. Examples include ginkgo, cycad, yew and conifer. A skilled artisan is aware of readily accessible databases that provide a comprehensive list of specific examples.

[0062] The term “levopimaradiene synthase” as used herein is defined as an enzyme which catalyzes the synthesis of levopimaradiene from geranylgeranyl

diphosphate through ionization of the allylic diphosphate moiety of labdadienyl pyrophosphate, followed by 1,4 hydrogen shift, methyl migration, and deprotonation.

## II. The Present Invention

[0063] Levopimaradiene synthase is useful to produce the ginkgolide precursor levopimaradiene. Potential levopimaradiene production methods of the present invention include *in vitro* conversion of geranylgeranyl diphosphate (GGDP) and *in vivo* production (in *Ginkgo* or microorganisms) using biosynthetic GGDP at native levels or in organisms genetically modified to increase the effective amount of geranylgeranyl diphosphate levels. The increase in the effective amount of GGDP allows more substrate (*e.g.*, GGDP) to be available for conversion to levopimaradiene and other enzyme diterpene products without the host organism suffering adverse consequences of low (*i.e.*, below required levels) GGDP levels.

[0064] Levopimaradiene synthase overexpression in *Ginkgo* in a specific embodiment allows increased levels of more advanced ginkgolide precursors. In alternative embodiments, additional genes are incorporated for increased quantities of levopimaradiene synthase, thereby leading to increased quantities of levopimaradiene or a ginkgolide. Expression of levopimaradiene synthase, which preferably does not contain a plastidial targeting sequence (see, for example, Peters *et al.* (2000); Williams *et al.* (2000)), in organisms that express genes encoding enzymes to metabolize GGDP, whether GGDP is exogenously provided or produced *de novo*, provide production of ginkgolide or ginkgolide precursors. One such ginkgolide precursor is levopimaradiene.

[0065] Levopimaradiene synthase, which directs the first committed step in ginkgolide biosynthesis, was cloned and characterized to ultimately isolate and functionally express genes involved in ginkgolide biosynthesis. This gene is essential to overproduction of ginkgolide using genetically modified organisms. A skilled artisan is aware that if the synthase exhibits low solubility and expression in *Escherichia coli*, *Saccharomyces cerevisiae* or other expression hosts, alternative strains and/or gene truncations are employed.

[0066] *Ginkgo biloba* levopimaradiene synthase is a cytosolically-synthesized plastid protein containing an N-terminal sequence that directs translocation of the levopimaradiene to specific plastidial compartments. The signal sequence is then excised

by a specific protease, yielding a mature levopimaradiene synthase. The optimal truncation site is determined through, for example, expression studies of the full-length gene and truncated versions, as described herein. The present invention contemplates a levopimaradiene synthase nucleic acid sequence and amino acid sequence that contains a deletion in the N-terminal sequence.

[0067] A skilled artisan is aware of standard means in the art to identify other levopimaradiene synthase nucleic acid sequences or other nucleic acid sequences which encode gene products that are functionally interchangeable with levopimaradiene synthase, meaning catalyze production of a diterpene, for example by searching publicly available sequence repositories such as GenBank or commercially available sequence repositories that are readily available. The SEQ.ID.NO:1 nucleic acid sequence is the *Ginkgo biloba* levopimaradiene synthase nucleic acid sequence, which encodes the *Ginkgo biloba* levopimaradiene synthase amino acid sequence (SEQ.ID.NO:2). A GenBank search with SEQ.ID.NO:1, the *Ginkgo biloba* levopimaradiene synthase nucleic acid sequence, identifies the similar sequence *Abies grandis* abietadiene synthase U50768.1 (SEQ.ID.NO:3) that encodes AAB05407 (SEQ.ID.NO:4), which is also in the scope of the present invention. A skilled artisan is aware of other standard methods to clone sequences, such as by library screening through hybridization to similar sequences.

[0068] Standard methods and reagents in the field of yeast molecular genetics, particularly regarding *Saccharomyces cerevisiae*, are well known in the art. References for such methods include Methods in Yeast Genetics, 2000 Edition: A Cold Spring Harbor Laboratory Course Manual (Burke *et al.*, 2000) and Current Protocols in Molecular Biology, Chapter 13 (Ausubel *et al.*, 1994), both incorporated by reference herein. A skilled artisan is aware that the *Saccharomyces* species of choice is *S. cerevisiae*, although there are other known species of the genus *Saccharomyces* including *S. italicus*, *S. oviformis*, *S. capensis*, *S. chevalieri*, *S. douglasii*, *S. paradoxus*, *S. cariocanus*, *S. kudriavzevii*, *S. mikatae*, *S. bayanus*, and *S. pastorianus*.

### III. Ginkgolides

[0069] A ginkgolide is a diterpenoid from the *Ginkgo biloba* plant. Examples include the following naturally occurring ginkgolides Ginkgolide A, Ginkgolide B, Ginkgolide C, Ginkgolide M, Ginkgolide J, in addition to other derivatives such as a

substituent(s) effecting solubility but not catalytic activity. A skilled artisan is aware of such moieties and methods to determine effects such as a desired solubility, electronic interaction, coordination and the other such properties without compromising biological activity. Preferable ginkgolides which are generated with the methods and compositions of the present invention include: Ginkgolide A and Ginkgolide B.

[0070] FIG. 3 demonstrates a generic ginkgolide structure with non-limiting examples of substituents for R<sub>1</sub>, R<sub>2</sub>, R<sub>3</sub> and R<sub>4</sub> given in the chart.

#### IV. Nucleic Acid-Based Expression Systems

##### A. Vectors

[0071] The term "vector" is used to refer to a carrier nucleic acid molecule into which a nucleic acid sequence is inserted for introduction into a cell where it is replicated. A nucleic acid sequence is in one instance "exogenous," which means that it is foreign to the cell into which the vector is being introduced or that the sequence is homologous to a sequence in the cell but in a position within the host cell nucleic acid in which the sequence is ordinarily not found. Vectors include plasmids, cosmids, viruses (bacteriophage, animal viruses, and plant viruses), and artificial chromosomes (e.g., YACs). One of skill in the art would be well equipped to construct a vector through standard recombinant techniques, which are described in Maniatis *et al.*, 1988 and Ausubel *et al.*, 1994, both incorporated herein by reference.

[0072] The term "expression vector" refers to a vector containing a nucleic acid sequence coding for at least part of a gene product capable of being transcribed. In some cases, RNA molecules are then translated into a protein, polypeptide, or peptide. In other cases, these sequences are not translated, for example, in the production of antisense molecules or ribozymes. Expression vectors, in one instance, contain a variety of "control sequences," which refer to nucleic acid sequences necessary for the transcription and possibly translation of an operably linked coding sequence in a particular host organism. In addition to control sequences that govern transcription and translation, vectors and expression vectors may contain nucleic acid sequences that serve other functions as well and are described *infra*.



## 1. Promoters and Enhancers

[0073] A “promoter” is a control sequence that is a region of a nucleic acid sequence at which initiation and rate of transcription are controlled. It may contain genetic elements at which regulatory proteins and molecules may bind such as RNA polymerase and other transcription factors. The phrases “operatively positioned,” “operatively linked,” “under control,” “under control of a promoter operable in” and “under transcriptional control” mean that a promoter is in a correct functional location and/or orientation in relation to a nucleic acid sequence to control transcriptional initiation and/or expression of that sequence. A promoter may or may not be used in conjunction with an “enhancer,” which refers to a *cis*-acting regulatory sequence involved in the transcriptional activation of a nucleic acid sequence.

[0074] A promoter may be one naturally associated with a gene or sequence, as may be obtained by isolating the 5' non-coding sequences located upstream of the coding segment and/or exon. Such a promoter is referred to as “endogenous.” Similarly, an enhancer may be one naturally associated with a nucleic acid sequence, located either downstream or upstream of that sequence. Alternatively, certain advantages are gained by positioning the coding nucleic acid segment under the control of a recombinant or heterologous promoter, which refers to a promoter that is not normally associated with a nucleic acid sequence in its natural environment. A recombinant or heterologous enhancer refers also to an enhancer not normally associated with a nucleic acid sequence in its natural environment. Such promoters or enhancers may include promoters or enhancers of other genes, and promoters or enhancers isolated from any other prokaryotic, viral, or eukaryotic cell, and promoters or enhancers not “naturally occurring,” *i.e.*, containing different elements of different transcriptional regulatory regions, and/or mutations that alter expression. In addition to producing nucleic acid sequences of promoters and enhancers synthetically, sequences may be produced using recombinant cloning and/or nucleic acid amplification technology, including PCR™, in connection with the compositions disclosed herein (see U.S. Patent 4,683,202, U.S. Patent 5,928,906, each incorporated herein by reference). Furthermore, it is contemplated the control sequences that direct transcription and/or expression of sequences within non-nuclear organelles such as mitochondria, chloroplasts, and the like, is employed as well.

[0075] Naturally, it is important to employ a promoter and/or enhancer that effectively directs the expression of the DNA segment in the cell type, organelle, and organism chosen for expression. Those of skill in the art of molecular biology generally know the use of promoters, enhancers, and cell type combinations for protein expression, for example, see Sambrook *et al.* (1989), incorporated herein by reference. The promoters employed may be constitutive, tissue-specific, inducible, and/or useful under the appropriate conditions to direct high level expression of the introduced DNA segment, such as is advantageous in the large-scale production of recombinant proteins and/or peptides. The promoter may be heterologous or endogenous.

[0076] Table 1 lists several elements/promoters that may be employed, in the context of the present invention, to regulate the expression of a gene. This list is not intended to be exhaustive of all the possible elements involved in the promotion of expression but, merely, to be exemplary thereof. Table 2 provides examples of inducible elements, which are regions of a nucleic acid sequence that is activated in response to a specific stimulus.

TABLE 1	
Promoter and/or Enhancer	
Promoter/Enhancer	References
Immunoglobulin Heavy Chain	Banerji <i>et al.</i> , 1983; Gilles <i>et al.</i> , 1983; Grosschedl <i>et al.</i> , 1985; Atchinson <i>et al.</i> , 1986, 1987; Imler <i>et al.</i> , 1987; Weinberger <i>et al.</i> , 1984; Kiledjian <i>et al.</i> , 1988; Porton <i>et al.</i> , 1990
Immunoglobulin Light Chain	Queen <i>et al.</i> , 1983; Picard <i>et al.</i> , 1984
T-Cell Receptor	Luria <i>et al.</i> , 1987; Winoto <i>et al.</i> , 1989; Redondo <i>et al.</i> , 1990
HLA DQ a and/or DQ $\beta$	Sullivan <i>et al.</i> , 1987
$\beta$ -Interferon	Goodbourn <i>et al.</i> , 1986; Fujita <i>et al.</i> , 1987; Goodbourn <i>et al.</i> , 1988
Interleukin-2	Greene <i>et al.</i> , 1989
Interleukin-2 Receptor	Greene <i>et al.</i> , 1989; Lin <i>et al.</i> , 1990
MHC Class II 5	Koch <i>et al.</i> , 1989
MHC Class II HLA-DRa	Sherman <i>et al.</i> , 1989
$\beta$ -Actin	Kawamoto <i>et al.</i> , 1988; Ng <i>et al.</i> , 1989

**TABLE 1**  
Promoter and/or Enhancer

Promoter/Enhancer	References
Muscle Creatine Kinase (MCK)	Jaynes <i>et al.</i> , 1988; Horlick <i>et al.</i> , 1989; Johnson <i>et al.</i> , 1989
Prealbumin (Transthyretin)	Costa <i>et al.</i> , 1988
Elastase I	Omitz <i>et al.</i> , 1987
Metallothionein (MTII)	Karin <i>et al.</i> , 1987; Culotta <i>et al.</i> , 1989
Collagenase	Pinkert <i>et al.</i> , 1987; Angel <i>et al.</i> , 1987
Albumin	Pinkert <i>et al.</i> , 1987; Tronche <i>et al.</i> , 1989, 1990
$\alpha$ -Fetoprotein	Godbout <i>et al.</i> , 1988; Campere <i>et al.</i> , 1989
t-Globin	Bodine <i>et al.</i> , 1987; Perez-Stable <i>et al.</i> , 1990
$\beta$ -Globin	Trudel <i>et al.</i> , 1987
c-fos	Cohen <i>et al.</i> , 1987
c-HA-ras	Triesman, 1986; Deschamps <i>et al.</i> , 1985
Insulin	Edlund <i>et al.</i> , 1985
Neural Cell Adhesion Molecule (NCAM)	Hirsh <i>et al.</i> , 1990
$\alpha_1$ -Antitrypsin	Latimer <i>et al.</i> , 1990
H2B (TH2B) Histone	Hwang <i>et al.</i> , 1990
Mouse and/or Type I Collagen	Ripe <i>et al.</i> , 1989
Glucose-Regulated Proteins (GRP94 and GRP78)	Chang <i>et al.</i> , 1989
Rat Growth Hormone	Larsen <i>et al.</i> , 1986
Human Serum Amyloid A (SAA)	Edbrooke <i>et al.</i> , 1989
Troponin I (TN I)	Yutzey <i>et al.</i> , 1989
Platelet-Derived Growth Factor (PDGF)	Pech <i>et al.</i> , 1989
Duchenne Muscular Dystrophy	Klamut <i>et al.</i> , 1990
SV40	Banerji <i>et al.</i> , 1981; Moreau <i>et al.</i> , 1981; Sleight <i>et al.</i> , 1985; Firak <i>et al.</i> , 1986; Herr <i>et al.</i> , 1986; Imbra <i>et al.</i> , 1986; Kadesch <i>et al.</i> , 1986; Wang <i>et al.</i> , 1986; Ondek <i>et al.</i> , 1987; Kuhl <i>et al.</i> , 1987; Schaffner <i>et al.</i> , 1988

**TABLE 1**  
Promoter and/or Enhancer

Promoter/Enhancer	References
Polyoma	Swartzendruber <i>et al.</i> , 1975; Vasseur <i>et al.</i> , 1980; Katinka <i>et al.</i> , 1980, 1981; Tyndell <i>et al.</i> , 1981; Dandolo <i>et al.</i> , 1983; de Villiers <i>et al.</i> , 1984; Hen <i>et al.</i> , 1986; Satake <i>et al.</i> , 1988; Campbell and/or Villarreal, 1988
Retroviruses	Kriegler <i>et al.</i> , 1982, 1983; Levinson <i>et al.</i> , 1982; Kriegler <i>et al.</i> , 1983, 1984a, b, 1988; Bosze <i>et al.</i> , 1986; Miksicek <i>et al.</i> , 1986; Celander <i>et al.</i> , 1987; Thiesen <i>et al.</i> , 1988; Celander <i>et al.</i> , 1988; Chol <i>et al.</i> , 1988; Reisman <i>et al.</i> , 1989
Papilloma Virus	Campo <i>et al.</i> , 1983; Lusky <i>et al.</i> , 1983; Spandidos and/or Wilkie, 1983; Spalholz <i>et al.</i> , 1985; Lusky <i>et al.</i> , 1986; Cripe <i>et al.</i> , 1987; Gloss <i>et al.</i> , 1987; Hirochika <i>et al.</i> , 1987; Stephens <i>et al.</i> , 1987; Glue <i>et al.</i> , 1988
Hepatitis B Virus	Bulla <i>et al.</i> , 1986; Jameel <i>et al.</i> , 1986; Shaul <i>et al.</i> , 1987; Spandau <i>et al.</i> , 1988; Vannice <i>et al.</i> , 1988
Human Immunodeficiency Virus	Muesing <i>et al.</i> , 1987; Hauber <i>et al.</i> , 1988; Jakobovits <i>et al.</i> , 1988; Feng <i>et al.</i> , 1988; Takebe <i>et al.</i> , 1988; Rosen <i>et al.</i> , 1988; Berkhout <i>et al.</i> , 1989; Laspia <i>et al.</i> , 1989; Sharp <i>et al.</i> , 1989; Braddock <i>et al.</i> , 1989
Cytomegalovirus (CMV)	Weber <i>et al.</i> , 1984; Boshart <i>et al.</i> , 1985; Foecking <i>et al.</i> , 1986
Gibbon Ape Leukemia Virus	Holbrook <i>et al.</i> , 1987; Quinn <i>et al.</i> , 1989

**TABLE 2**  
Inducible Elements

Element	Inducer	References
MT II	Phorbol Ester (TFA) Heavy metals	Palmiter <i>et al.</i> , 1982; Haslinger <i>et al.</i> , 1985; Searle <i>et al.</i> , 1985; Stuart <i>et al.</i> , 1985; Imagawa <i>et al.</i> , 1987, Karin <i>et al.</i> , 1987; Angel <i>et al.</i> , 1987b; McNeall <i>et al.</i> , 1989
MMTV (mouse mammary tumor virus)	Glucocorticoids	Huang <i>et al.</i> , 1981; Lee <i>et al.</i> , 1981; Majors <i>et al.</i> , 1983; Chandler <i>et al.</i> , 1983; Lee <i>et al.</i> , 1984; Ponta <i>et al.</i> , 1985; Sakai <i>et al.</i> , 1988
$\beta$ -Interferon	poly(rI)x poly(rc)	Tavernier <i>et al.</i> , 1983
Adenovirus 5 E2	E1A	Imperiale <i>et al.</i> , 1984
Collagenase	Phorbol Ester (TPA)	Angel <i>et al.</i> , 1987a
Stromelysin	Phorbol Ester (TPA)	Angel <i>et al.</i> , 1987b
SV40	Phorbol Ester (TPA)	Angel <i>et al.</i> , 1987b
Murine MX Gene	Interferon, Newcastle Disease Virus	Hug <i>et al.</i> , 1988
GRP78 Gene	A23187	Resendez <i>et al.</i> , 1988
$\alpha$ -2-Macroglobulin	IL-6	Kunz <i>et al.</i> , 1989
Vimentin	Serum	Rittling <i>et al.</i> , 1989
MHC Class I Gene H-2kb	Interferon	Blonar <i>et al.</i> , 1989
HSP70	E1A, SV40 Large T Antigen	Taylor <i>et al.</i> , 1989, 1990a, 1990b
Proliferin	Phorbol Ester-TPA	Mordacq <i>et al.</i> , 1989
Tumor Necrosis Factor	PMA	Hensel <i>et al.</i> , 1989
Thyroid Stimulating Hormone $\alpha$ Gene	Thyroid Hormone	Chatterjee <i>et al.</i> , 1989

[0077] The identity of tissue-specific promoters or elements, as well as assays to characterize their activity, is well known to those of skill in the art. Examples of such

regions include the human LIMK2 gene (Nomoto *et al.* 1999), the somatostatin receptor 2 gene (Kraus *et al.*, 1998), murine epididymal retinoic acid-binding gene (Lareyre *et al.*, 1999), human CD4 (Zhao-Emonet *et al.*, 1998), mouse alpha2 (XI) collagen (Tsumaki, *et al.*, 1998), D1A dopamine receptor gene (Lee, *et al.*, 1997), insulin-like growth factor II (Wu *et al.*, 1997), human platelet endothelial cell adhesion molecule-1 (Almendo *et al.*, 1996).

## **2. Initiation Signals and Internal Ribosome Binding Sites**

[0078] A specific initiation signal also may be required for efficient translation of coding sequences. These signals include the ATG initiation codon or adjacent sequences. Exogenous translational control signals, including the ATG initiation codon, may need to be provided. One of ordinary skill in the art would readily be capable of determining this and providing the necessary signals. It is well known that the initiation codon must be "in-frame" with the reading frame of the desired coding sequence to ensure translation of the entire insert. The exogenous translational control signals and initiation codons are either natural or synthetic. The efficiency of expression may be enhanced by the inclusion of appropriate transcription enhancer elements.

[0079] In certain embodiments of the invention, the use of internal ribosome entry sites (IRES) elements are used to create multigene, or polycistronic, messages. IRES elements are able to bypass the ribosome scanning model of 5' methylated Cap dependent translation and begin translation at internal sites (Pelletier and Sonenberg, 1988). IRES elements from two members of the picornavirus family (polio and encephalomyocarditis) have been described (Pelletier and Sonenberg, 1988), as well as IRES from a mammalian message (Macejak and Sarnow, 1991). IRES elements are, in one instance, linked to heterologous open reading frames. Multiple open reading frames are transcribed together, each separated by an IRES, creating polycistronic messages. By virtue of the IRES element, each open reading frame is accessible to ribosomes for efficient translation. Multiple genes are efficiently expressed using a single promoter/enhancer to transcribe a single message (see U.S. Patent 5,925,565 and 5,935,819, herein incorporated by reference).

### 3. Multiple Cloning Sites

[0080] Vectors include, in some instances, a multiple cloning site (MCS), which is a nucleic acid region that contains multiple restriction enzyme sites, any of which are used in conjunction with standard recombinant technology to digest the vector. (See Carbonelli *et al.*, 1999, Levenson *et al.*, 1998, and Cocea, 1997, incorporated herein by reference.) "Restriction enzyme digestion" refers to catalytic cleavage of a nucleic acid molecule with an enzyme that functions only at specific locations in a nucleic acid molecule. Many of these restriction enzymes are commercially available. Use of such enzymes is widely understood by those of skill in the art. Frequently, a vector is linearized or fragmented using a restriction enzyme that cuts within the MCS to enable exogenous sequences to be ligated to the vector. "Ligation" refers to the process of forming phosphodiester bonds between two nucleic acid fragments, which may or may not be contiguous with each other. Techniques involving restriction enzymes and ligation reactions are well known to those of skill in the art of recombinant technology.

### 4. Splicing Sites

[0081] Most transcribed eukaryotic RNA molecules undergo RNA splicing to remove introns from the primary transcripts. Vectors containing genomic eukaryotic sequences may require donor and/or acceptor splicing sites to ensure proper processing of the transcript for protein expression. (See Chandler *et al.*, 1997, herein incorporated by reference.)

### 5. Polyadenylation Signals

[0082] In expression, one typically includes a polyadenylation signal to effect proper polyadenylation of the transcript. The nature of the polyadenylation signal is not believed to be crucial to the successful practice of the invention, and/or any such sequence may be employed. Preferred embodiments include the SV40 polyadenylation signal and/or the bovine growth hormone polyadenylation signal, convenient and/or known to function well in various target cells. Also contemplated as an element of the expression cassette is a transcriptional termination site. These elements serve to enhance message levels and/or to minimize read through from the cassette into other sequences.

## 6. Origins of Replication

[0083] In order to propagate a vector in a host cell, it may contain one or more origins of replication sites (often termed "ori"), which is a specific nucleic acid sequence at which replication is initiated. Alternatively an autonomously replicating sequence (ARS) is employed if the host cell is yeast.

## 7. Selectable and Screenable Markers

[0084] In certain embodiments of the invention, the cells contain nucleic acid construct of the present invention, a cell may be identified *in vitro* or *in vivo* by including a marker in the expression vector. Such markers would confer an identifiable change to the cell permitting easy identification of cells containing the expression vector. Generally, a selectable marker is one that confers a property that allows for selection. A positive selectable marker is one in which the presence of the marker allows for its selection, while a negative selectable marker is one in which its presence prevents its selection. An example of a positive selectable marker is a drug resistance marker.

[0085] Usually the inclusion of a drug selection marker aids in the cloning and identification of transformants, for example, genes that confer resistance to neomycin, puromycin, hygromycin, DHFR, GPT, zeocin and histidinol are useful selectable markers. In addition to markers conferring a phenotype that allows for the discrimination of transformants based on the implementation of conditions, other types of markers including screenable markers such as GFP, whose basis is colorimetric analysis, are also contemplated. Alternatively, screenable enzymes such as herpes simplex virus thymidine kinase (*tk*) or chloramphenicol acetyltransferase (CAT) may be utilized. One of skill in the art would also know how to employ immunologic markers, possibly in conjunction with FACS analysis. The marker used is not believed to be important, so long as it is capable of being expressed simultaneously with the nucleic acid encoding a gene product, such as a levopimaradiene synthase. Further examples of selectable and screenable markers are well known to one of skill in the art, such as amino acid markers including, but not limited to, uracil, leucine, tryptophan and histidine biosynthetic genes. A host that is auxotrophic for the amino acid biosynthetic gene used as a selectable marker allows ready screening for transformer cells comprising the nucleic acid sequence of interest.



## B. Host Cells

[0086] As used herein, the terms "cell," "cell line," and "cell culture" may be used interchangeably. All of these term also include their progeny, which is any and all subsequent generations. It is understood that all progeny may not be identical due to deliberate or inadvertent mutations. In the context of expressing a heterologous nucleic acid sequence, "host cell" refers to a prokaryotic or eukaryotic cell, and it includes any transformable organism that is capable of replicating a vector and/or expressing a heterologous gene encoded by a vector. A host cell is, in most instances, used as a recipient for vectors. A host cell may be "transfected" or "transformed," which refers to a process by which exogenous nucleic acid is transferred or introduced into the host cell. A transformed cell includes the primary subject cell and its progeny.

[0087] Host cells may be derived from prokaryotes or eukaryotes, depending upon whether the desired result is replication of the vector or expression of part or all of the vector-encoded nucleic acid sequences. Numerous cell lines and cultures are available for use as a host cell, and they are obtained through, for example, the American Type Culture Collection (ATCC), which is an organization that serves as an archive for living cultures and genetic materials. An appropriate host is determined by one of skill in the art based on the vector backbone and the desired result. A plasmid or cosmid, for example, is introduced into a prokaryote host cell for replication of many vectors. Bacterial cells used as host cells for vector replication and/or expression include DH5 $\alpha$ , JM109, and KC8, as well as a number of commercially available bacterial hosts such as SURE<sup>®</sup> Competent Cells and SOLOPACK<sup>™</sup> Gold Cells (STRATAGENE<sup>®</sup>, La Jolla). Alternatively, bacterial cells such as *E. coli* LE392 are used as host cells for phage viruses.

[0088] Examples of eukaryotic host cells for replication and/or expression of a vector include HeLa, NIH3T3, Jurkat, 293, Cos, CHO, Saos, and PC12. Many host cells from various cell types and organisms are available and would be known to one of skill in the art. Similarly, a viral vector may be used in conjunction with either a eukaryotic or prokaryotic host cell, particularly one that is permissive for replication or expression of the vector.

[0089] Some vectors may employ control sequences that allow it to be replicated and/or expressed in both prokaryotic and eukaryotic cells. One of skill in the art would further understand the conditions under which to incubate all of the above described host

cells to maintain them and to permit replication of a vector. Also understood and known are techniques and conditions that would allow large-scale production of vectors, as well as production of the nucleic acids encoded by vectors and their cognate polypeptides, proteins, or peptides.

[0090] Another such host cell is a cell that accumulates an increase in the amount of geranylgeranyl diphosphate that is biosynthesized *de novo*. An example of such a microorganism is described in co-pending application "Diterpene-producing unicellular organism", filed on the same day as the instant application. The increase in the amount of substrate for levopimaradiene synthase (e.g., geranylgeranyl diphosphate) allows a proportional increase in levopimaradiene production.

### C. Expression Systems

[0091] Numerous expression systems exist that comprise at least a part or all of the compositions discussed above. Prokaryote- and/or eukaryote-based systems are employed for use with the present invention to produce nucleic acid sequences, or their cognate polypeptides, proteins and peptides. Many such systems are commercially and widely available.

[0092] The insect cell/baculovirus system produce a high level of protein expression of a heterologous nucleic acid segment, such as described in U.S. Patent No. 5,871,986, 4,879,236, both herein incorporated by reference, and which are bought, for example, under the name MAXBAC<sup>®</sup> 2.0 from INVITROGEN<sup>®</sup> and BACPACK<sup>™</sup> BACULOVIRUS EXPRESSION SYSTEM FROM CLONTECH<sup>®</sup>.

[0093] Other examples of expression systems include STRATAGENE<sup>®</sup>'s COMPLETE CONTROL<sup>™</sup> Inducible Mammalian Expression System, which involves a synthetic ecdysone-inducible receptor, or its *E. coli* pET Bacterial Expression System. Another example of an inducible expression system is available from INVITROGEN<sup>®</sup>, which carries the T-REX<sup>™</sup> (tetracycline-regulated expression) System, an inducible mammalian expression system that uses the full-length CMV promoter. INVITROGEN<sup>®</sup> also provides a yeast expression system called the *Pichia methanolica* Expression System, which is designed for high-level production of recombinant proteins in the methylotrophic yeast *Pichia methanolica*. One of skill in the art would know how to express a vector, such as

an expression construct, to produce a nucleic acid sequence or its cognate polypeptide, protein, or peptide.

## **V. Nucleic Acid Detection**

[0094] In addition to their use in directing the expression of levopimaradiene synthase proteins, polypeptides and/or peptides, the nucleic acid sequences disclosed herein have a variety of other uses. For example, they have utility as probes or primers for embodiments involving nucleic acid hybridization.

### **A. Hybridization**

[0095] The use of a probe or primer of between 13 and 100 nucleotides, preferably between 17 and 100 nucleotides in length, or in some aspects of the invention up to 1-2 kilobases or more in length, allows the formation of a duplex molecule that is both stable and selective. Molecules having complementary sequences over contiguous stretches greater than 20 bases in length are generally preferred, to increase stability and/or selectivity of the hybrid molecules obtained. One generally prefers to design nucleic acid molecules for hybridization having one or more complementary sequences of 20 to 30 nucleotides, or even longer where desired. Such fragments may be readily prepared, for example, by directly synthesizing the fragment by chemical means or by introducing selected sequences into recombinant vectors for recombinant production.

[0096] Accordingly, the nucleotide sequences of the invention may be used for their ability to selectively form duplex molecules with complementary stretches of DNAs and/or RNAs or to provide primers for amplification of DNA or RNA from samples. Depending on the application envisioned, one would desire to employ varying conditions of hybridization to achieve varying degrees of selectivity of the probe or primers for the target sequence.

[0097] For applications requiring high selectivity, one typically desires to employ relatively high stringency conditions to form the hybrids. For example, relatively low salt and/or high temperature conditions, such as provided by about 0.02 M to about 0.10 M NaCl at temperatures of about 50°C to about 70°C. Such high stringency conditions tolerate little, if any, mismatch between the probe or primers and the template or target strand and would be particularly suitable for isolating specific genes or for detecting specific mRNA transcripts. It is generally appreciated that conditions are rendered more stringent by the addition of increasing amounts of formamide.

10041007-010702

[0098] For certain applications, for example, site-directed mutagenesis, it is appreciated that lower stringency conditions are preferred. Under these conditions, hybridization may occur even though the sequences of the hybridizing strands are not perfectly complementary, but are mismatched at one or more positions. Conditions may be rendered less stringent by increasing salt concentration and/or decreasing temperature. For example, a medium stringency condition could be provided by about 0.1 to 0.25 M NaCl at temperatures of about 37°C to about 55°C, while a low stringency condition could be provided by about 0.15 M to about 0.9 M salt, at temperatures ranging from about 20°C to about 55°C. Hybridization conditions are readily manipulated depending on the desired results.

[0099] In other embodiments, hybridization may be achieved under conditions of, for example, 50 mM Tris-HCl (pH 8.3), 75 mM KCl, 3 mM MgCl<sub>2</sub>, 1.0 mM dithiothreitol, at temperatures between approximately 20°C to about 37°C. Other hybridization conditions utilized could include approximately 10 mM Tris-HCl (pH 8.3), 50 mM KCl, 1.5 mM MgCl<sub>2</sub>, at temperatures ranging from approximately 40°C to about 72°C.

[0100] In certain embodiments, it is advantageous to employ nucleic acids of defined sequences of the present invention in combination with an appropriate means, such as a label, for determining hybridization. A wide variety of appropriate indicator means are known in the art, including fluorescent, radioactive, enzymatic or other ligands, such as avidin/biotin, which are capable of being detected. In preferred embodiments, one may desire to employ a fluorescent label or an enzyme tag such as urease, alkaline phosphatase or peroxidase, instead of radioactive or other environmentally undesirable reagents. In the case of enzyme tags, colorimetric indicator substrates are known that are employed to provide a detection means that is visibly or spectrophotometrically detectable, to identify specific hybridization with complementary nucleic acid containing samples.

[0101] In general, it is envisioned that the probes or primers described herein are useful as reagents in solution hybridization, as in PCR™, for detection of expression of corresponding genes, as well as in embodiments employing a solid phase. In embodiments involving a solid phase, the test DNA (or RNA) is adsorbed or otherwise affixed to a selected matrix or surface. This fixed, single-stranded nucleic acid is then subjected to hybridization with selected probes under desired conditions. The conditions

selected depend on the particular circumstances (depending, for example, on the G+C content, type of target nucleic acid, source of nucleic acid, size of hybridization probe, etc.). Optimization of hybridization conditions for the particular application of interest is well known to those of skill in the art. After washing of the hybridized molecules to remove non-specifically bound probe molecules, hybridization is detected, and/or quantified, by determining the amount of bound label. Representative solid phase hybridization methods are disclosed in U.S. Patent Nos. 5,843,663, 5,900,481 and 5,919,626. Other methods of hybridization that may be used in the practice of the present invention are disclosed in U.S. Patent Nos. 5,849,481, 5,849,486 and 5,851,772. The relevant portions of these and other references identified in this section of the Specification are incorporated herein by reference.

#### **B. Amplification of Nucleic Acids**

[00102] Nucleic acids used as a template for amplification may be isolated from cells, tissues or other samples according to standard methodologies (Sambrook *et al.*, 1989). In certain embodiments, analysis is performed on whole cell or tissue homogenates or biological fluid samples without substantial purification of the template nucleic acid. The nucleic acid may be genomic DNA or fractionated or whole cell RNA. Where RNA is used, it may be desired to first convert the RNA to a complementary DNA.

[00103] The term "primer," as used herein, is meant to encompass any nucleic acid that is capable of priming the synthesis of a nascent nucleic acid in a template-dependent process. Typically, primers are oligonucleotides from ten to twenty and/or thirty base pairs in length, but longer sequences are also contemplated. Primers are provided in double-stranded and/or single-stranded form, although the single-stranded form is preferred.

[00104] Pairs of primers designed to selectively hybridize to nucleic acids corresponding to levodipyrrole synthase are contacted with the template nucleic acid under conditions that permit selective hybridization. Depending upon the desired application, high stringency hybridization conditions are selected that only allow hybridization to sequences that are completely complementary to the primers. In other embodiments, hybridization occurs under reduced stringency to allow for amplification of nucleic acids contain one or more mismatches with the primer sequences. Once hybridized, the template-primer complex is contacted with one or more enzymes that

facilitate template-dependent nucleic acid synthesis. Multiple rounds of amplification, also referred to as "cycles," are conducted until a sufficient amount of amplification product is produced.

[00105] The amplification product is detected or quantified. In certain applications, the detection is performed by visual means. Alternatively, the detection involves indirect identification of the product *via* chemiluminescence, radioactive scintigraphy of incorporated radiolabel or fluorescent label or even *via* a system using electrical and/or thermal impulse signals (Affymax technology; Bellus, 1994).

[00106] A number of template dependent processes are available to amplify the oligonucleotide sequences present in a given template sample. One of the best known amplification methods is the polymerase chain reaction (referred to as PCR<sup>TM</sup>) which is described in detail in U.S. Patent Nos. 4,683,195, 4,683,202 and 4,800,159, and in Innis *et al.*, 1990, each of which is incorporated herein by reference in their entirety.

[00107] A reverse transcriptase PCR<sup>TM</sup> amplification procedure is performed to quantify the amount of mRNA amplified. Methods of reverse transcribing RNA into cDNA are well known and described in Sambrook *et al.*, 1989. Alternative methods for reverse transcription utilize thermostable DNA polymerases. These methods are described in WO 90/07641. Polymerase chain reaction methodologies are well known in the art. Representative methods of RT-PCR are described in U.S. Patent No. 5,882,864.

[00108] Another method for amplification is ligase chain reaction ("LCR"), disclosed in European Application No. 320 308, incorporated herein by reference in its entirety. U.S. Patent 4,883,750 describes a method similar to LCR for binding probe pairs to a target sequence. A method based on PCR<sup>TM</sup> and oligonucleotide ligase assay (OLA), disclosed in U.S. Patent 5,912,148, may also be used.

[00109] Alternative methods for amplification of target nucleic acid sequences that are used in the practice of the present invention are disclosed in U.S. Patent Nos. 5,843,650, 5,846,709, 5,846,783, 5,849,546, 5,849,497, 5,849,547, 5,858,652, 5,866,366, 5,916,776, 5,922,574, 5,928,905, 5,928,906, 5,932,451, 5,935,825, 5,939,291 and 5,942,391, GB Application No. 2 202 328, and in PCT Application No. PCT/US89/01025, each of which is incorporated herein by reference in its entirety.

[00110] Qbeta Replicase, described in PCT Application No. PCT/US87/00880, is also used as an amplification method in the present invention. In this

10044007-010702

method, a replicative sequence of RNA that has a region complementary to that of a target is added to a sample in the presence of an RNA polymerase. The polymerase copies the replicative sequence which then are detected.

[00111] An isothermal amplification method, in which restriction endonucleases and ligases are used to achieve the amplification of target molecules that contain nucleotide 5'-[alpha-thio]-triphosphates in one strand of a restriction site may also be useful in the amplification of nucleic acids in the present invention (Walker *et al.*, 1992). Strand Displacement Amplification (SDA), disclosed in U.S. Patent No. 5,916,779, is another method of carrying out isothermal amplification of nucleic acids which involves multiple rounds of strand displacement and synthesis, *i.e.*, nick translation.

[00112] Other nucleic acid amplification procedures include transcription-based amplification systems (TAS), including nucleic acid sequence based amplification (NASBA) and 3SR (Kwoh *et al.*, 1989; Gingeras *et al.*, PCT Application WO 88/10315, incorporated herein by reference in their entirety). Davey *et al.*, European Application No. 329 822 disclose a nucleic acid amplification process involving cyclically synthesizing single-stranded RNA ("ssRNA"), ssDNA, and double-stranded DNA (dsDNA), which is used in accordance with the present invention.

[00113] Miller *et al.*, PCT Application WO 89/06700 (incorporated herein by reference in its entirety) disclose a nucleic acid sequence amplification scheme based on the hybridization of a promoter region/primer sequence to a target single-stranded DNA ("ssDNA") followed by transcription of many RNA copies of the sequence. This scheme is not cyclic, *i.e.*, new templates are not produced from the resultant RNA transcripts. Other amplification methods include "race" and "one-sided PCR" (Frohman, 1990; Ohara *et al.*, 1989).

### **C. Detection of Nucleic Acids**

[00114] Following any amplification, it may be desirable to separate the amplification product from the template and/or the excess primer. In one embodiment, amplification products are separated by agarose, agarose-acrylamide or polyacrylamide gel electrophoresis using standard methods (Sambrook *et al.*, 1989). Separated amplification products are cut out and eluted from the gel for further manipulation. Using

low melting point agarose gels, the separated band is removed by heating the gel, followed by extraction of the nucleic acid.

[00115] Separation of nucleic acids may also be effected by chromatographic techniques known in art. There are many kinds of chromatography which may be used in the practice of the present invention, including adsorption, partition, ion-exchange, hydroxylapatite, molecular sieve, reverse-phase, column, paper, thin-layer, and gas chromatography as well as HPLC.

[00116] In certain embodiments, the amplification products are visualized. A typical visualization method involves staining of a gel with ethidium bromide and visualization of bands under UV light. Alternatively, if the amplification products are integrally labeled with radio- or fluorometrically-labeled nucleotides, the separated amplification products are exposed to x-ray film or visualized under the appropriate excitatory spectra.

[00117] In one embodiment, following separation of amplification products, a labeled nucleic acid probe is brought into contact with the amplified marker sequence. The probe preferably is conjugated to a chromophore but may be radiolabeled. In another embodiment, the probe is conjugated to a binding partner, such as an antibody or biotin, or another binding partner carrying a detectable moiety.

[00118] In particular embodiments, detection is by Southern blotting and hybridization with a labeled probe. The techniques involved in Southern blotting are well known to those of skill in the art. See Sambrook *et al.*, 1989. One example of the foregoing is described in U.S. Patent No. 5,279,721, incorporated by reference herein, which discloses an apparatus and method for the automated electrophoresis and transfer of nucleic acids. The apparatus permits electrophoresis and blotting without external manipulation of the gel and is ideally suited to carrying out methods according to the present invention.

[00119] Other methods of nucleic acid detection that are used in the practice of the instant invention are disclosed in U.S. Patent Nos. 5,840,873, 5,843,640, 5,843,651, 5,846,708, 5,846,717, 5,846,726, 5,846,729, 5,849,487, 5,853,990, 5,853,992, 5,853,993, 5,856,092, 5,861,244, 5,863,732, 5,863,753, 5,866,331, 5,905,024, 5,910,407, 5,912,124, 5,912,145, 5,919,630, 5,925,517, 5,928,862, 5,928,869, 5,929,227, 5,932,413 and 5,935,791, each of which is incorporated herein by reference.



#### **D. Other Assays**

**[00120]** Other methods for genetic screening are used within the scope of the present invention, for example, to detect mutations in genomic DNA, cDNA and/or RNA samples. Methods used to detect point mutations include denaturing gradient gel electrophoresis ("DGGE"), restriction fragment length polymorphism analysis ("RFLP"), chemical or enzymatic cleavage methods, direct sequencing of target regions amplified by PCR<sup>TM</sup> (see above), single-strand conformation polymorphism analysis ("SSCP") and other methods well known in the art.

**[00121]** One method of screening for point mutations is based on RNase cleavage of base pair mismatches in RNA/DNA or RNA/RNA heteroduplexes. As used herein, the term "mismatch" is defined as a region of one or more unpaired or mispaired nucleotides in a double-stranded RNA/RNA, RNA/DNA or DNA/DNA molecule. This definition thus includes mismatches due to insertion/deletion mutations, as well as single or multiple base point mutations.

**[00122]** U.S. Patent No. 4,946,773 describes an RNase A mismatch cleavage assay that involves annealing single-stranded DNA or RNA test samples to an RNA probe, and subsequent treatment of the nucleic acid duplexes with RNase A. For the detection of mismatches, the single-stranded products of the RNase A treatment, electrophoretically separated according to size, are compared to similarly treated control duplexes. Samples containing smaller fragments (cleavage products) not seen in the control duplex are scored as positive.

**[00123]** Other investigators have described the use of RNase I in mismatch assays. The use of RNase I for mismatch detection is described in literature from Promega Biotech. Promega markets a kit containing RNase I that is reported to cleave three out of four known mismatches. Others have described using the MutS protein or other DNA-repair enzymes for detection of single-base mismatches.

**[00124]** Alternative methods for detection of deletion, insertion or substitution mutations that are used in the practice of the present invention are disclosed in U.S. Patent Nos. 5,849,483, 5,851,770, 5,866,337, 5,925,525 and 5,928,870, each of which is incorporated herein by reference in its entirety.

## VI. Site-Directed Mutagenesis

[00125] Structure-guided site-specific mutagenesis (also called site-directed mutagenesis) represents a powerful tool for the dissection and engineering of protein-ligand interactions (Wells, 1996, Braisted *et al.*, 1996). The technique provides for the preparation and testing of sequence variants by introducing one or more nucleotide sequence changes into a selected DNA.

[00126] Site-specific mutagenesis uses specific oligonucleotide sequences which encode the DNA sequence of the desired mutation, as well as a sufficient number of adjacent, unmodified nucleotides. In this way, a primer sequence is provided with sufficient size and complexity to form a stable duplex on both sides of the deletion junction being traversed. A primer of about 17 to 25 nucleotides in length is preferred, with about 5 to 10 residues on both sides of the junction of the sequence being altered.

[00127] The technique typically employs a bacteriophage vector that exists in both a single-stranded and double-stranded form. Vectors useful in site-directed mutagenesis include vectors such as the M13 phage. These phage vectors are commercially available and their use is generally well known to those skilled in the art. Double-stranded plasmids are also routinely employed in site-directed mutagenesis, which eliminates the step of transferring the gene of interest from a phage to a plasmid.

[00128] In general, one first obtains a single-stranded vector, or melts two strands of a double-stranded vector, which includes within its sequence a DNA sequence encoding the desired protein or genetic element. An oligonucleotide primer bearing the desired mutated sequence, synthetically prepared, is then annealed with the single-stranded DNA preparation, taking into account the degree of mismatch when selecting hybridization conditions. The hybridized product is subjected to DNA polymerizing enzymes such as *E. coli* polymerase I (Klenow fragment) in order to complete the synthesis of the mutation-bearing strand. Thus, a heteroduplex is formed, wherein one strand encodes the original non-mutated sequence, and the second strand bears the desired mutation. This heteroduplex vector is then used to transform appropriate host cells, such as *E. coli* cells, and clones are selected that include recombinant vectors bearing the mutated sequence arrangement.

[00129] Comprehensive information on the functional significance and information content of a given residue of protein is best be obtained by saturation

mutagenesis in which all 19 amino acid substitutions are examined. The shortcoming of this approach is that the logistics of multiresidue saturation mutagenesis are daunting (Warren *et al.*, 1996, Brown *et al.*, 1996; Zeng *et al.*, 1996; Burton and Barbas, 1994; Yelton *et al.*, 1995; Jackson *et al.*, 1995; Short *et al.*, 1995; Wong *et al.*, 1996; Hilton *et al.*, 1996). Hundreds, and possibly even thousands, of site specific mutants must be studied. However, improved techniques make production and rapid screening of mutants much more straightforward. See also, U.S. Patents 5,798,208 and 5,830,650, for a description of "walk-through" mutagenesis.

[00130] Other methods of site-directed mutagenesis are disclosed in U.S. Patents 5,220,007; 5,284,760; 5,354,670; 5,366,878; 5,389,514; 5,635,377; and 5,789,166.

## VII. Levopimaradiene Synthase Nucleic Acids

### A. Nucleic Acids and Uses Thereof

[00131] Certain aspects of the present invention concern at least one levopimaradiene synthase nucleic acid. In certain aspects, the at least one levopimaradiene synthase nucleic acid comprises a wild-type or mutant levopimaradiene synthase nucleic acid. In particular aspects, the levopimaradiene synthase nucleic acid encodes for at least one transcribed nucleic acid. In certain aspects, the levopimaradiene synthase nucleic acid comprises at least one transcribed nucleic acid. In particular aspects, the levopimaradiene synthase nucleic acid encodes at least one levopimaradiene synthase protein, polypeptide or peptide, or biologically functional equivalent thereof. In other aspects, the levopimaradiene synthase nucleic acid comprises at least one nucleic acid segment of SEQ.ID.NO:1, or at least one biologically functional equivalent thereof, for example SEQ.ID.NO:32, SEQ.ID.NO:34, SEQ.ID.NO:36, or SEQ.ID.NO:38.

[00132] A skilled artisan is aware that a nucleic acid sequence of the present invention may be contained on an episome, such as a plasmid or other vector, or may be on a chromosome of an organism, or both.

[00133] The present invention also concerns the isolation or creation of at least one recombinant construct or at least one recombinant host cell through the application of recombinant nucleic acid technology known to those of skill in the art or as described herein. The recombinant construct or host cell may comprise at least one levopimaradiene synthase nucleic acid, and may express at least one levopimaradiene

synthase protein, polypeptide or peptide, or at least one biologically functional equivalent thereof.

[00134] As used herein "wild-type" refers to the naturally occurring sequence of a nucleic acid at a genetic locus in the genome of an organism, and sequences transcribed or translated from such a nucleic acid. Thus, the term "wild-type" also may refer to the amino acid sequence encoded by the nucleic acid. As a genetic locus may have more than one sequence or alleles in a population of individuals, the term "wild-type" encompasses all such naturally occurring alleles. As used herein the term "polymorphic" means that variation exists (*i.e.* two or more alleles exist) at a genetic locus in the individuals of a population. As used herein "mutant" refers to a change in the sequence of a nucleic acid or its encoded protein, polypeptide or peptide that is the result of the hand of man.

[00135] A nucleic acid may be made by any technique known to one of ordinary skill in the art. Non-limiting examples of synthetic nucleic acid, particularly a synthetic oligonucleotide, include a nucleic acid made by *in vitro* chemical synthesis using phosphotriester, phosphite or phosphoramidite chemistry and solid phase techniques such as described in EP 266,032, incorporated herein by reference, or *via* deoxynucleoside H-phosphonate intermediates as described by Froehler *et al.*, 1986, and U.S. Patent Serial No. 5,705,629, each incorporated herein by reference. A non-limiting example of enzymatically produced nucleic acid include one produced by enzymes in amplification reactions such as PCR<sup>TM</sup> (see for example, U.S. Patent 4,683,202 and U.S. Patent 4,682,195, each incorporated herein by reference), or the synthesis of oligonucleotides described in U.S. Patent No. 5,645,897, incorporated herein by reference. A non-limiting example of a biologically produced nucleic acid includes recombinant nucleic acid production in living cells, such as recombinant DNA vector production in bacteria (see for example, Sambrook *et al.* 1989, incorporated herein by reference).

[00136] A nucleic acid may be purified on polyacrylamide gels, cesium chloride centrifugation gradients, or by any other means known to one of ordinary skill in the art (see for example, Sambrook *et al.* 1989, incorporated herein by reference).

[00137] The term "nucleic acid" generally refers to at least one molecule or strand of DNA, RNA or a derivative or mimic thereof, comprising at least one nucleobase, such as, for example, a naturally occurring purine or pyrimidine base found in DNA (*e.g.*

adenine "A," guanine "G," thymine "T" and cytosine "C") or RNA (*e.g.* A, G, uracil "U" and C). The term "nucleic acid" encompass the terms "oligonucleotide" and "polynucleotide." The term "oligonucleotide" refers to at least one molecule of between about 3 and about 100 nucleobases in length. The term "polynucleotide" refers to at least one molecule of greater than about 100 nucleobases in length. These definitions generally refer to at least one single-stranded molecule, but in specific embodiments also encompass at least one additional strand that is partially, substantially or fully complementary to the at least one single-stranded molecule. Thus, a nucleic acid may encompass at least one double-stranded molecule or at least one triple-stranded molecule that comprises one or more complementary strand(s) or "complement(s)" of a particular sequence comprising a strand of the molecule. As used herein, a single stranded nucleic acid may be denoted by the prefix "ss", a double stranded nucleic acid by the prefix "ds", and a triple stranded nucleic acid by the prefix "ts."

[00138] Thus, the present invention also encompasses at least one nucleic acid that is complementary to a levopimaradiene synthase nucleic acid. In particular embodiments the invention encompasses at least one nucleic acid or nucleic acid segment complementary to the sequence set forth in SEQ.ID.NO:1, SEQ.ID.NO:32, SEQ.ID.NO:34, SEQ.ID.NO:36, and/or SEQ.ID.NO:38. Nucleic acid(s) that are "complementary" or "complement(s)" are those that are capable of base-pairing according to the standard Watson-Crick, Hoogsteen or reverse Hoogsteen binding complementarity rules. As used herein, the term "complementary" or "complement(s)" also refers to nucleic acid(s) that are substantially complementary, as may be assessed by the same nucleotide comparison set forth above. The term "substantially complementary" refers to a nucleic acid comprising at least one sequence of consecutive nucleobases, or semiconsecutive nucleobases if one or more nucleobase moieties are not present in the molecule, are capable of hybridizing to at least one nucleic acid strand or duplex even if less than all nucleobases do not base pair with a counterpart nucleobase. In certain embodiments, a "substantially complementary" nucleic acid contains at least one sequence in which about 70%, about 71%, about 72%, about 73%, about 74%, about 75%, about 76%, about 77%, about 77%, about 78%, about 79%, about 80%, about 81%, about 82%, about 83%, about 84%, about 85%, about 86%, about 87%, about 88%, about 89%, about 90%, about 91%, about 92%, about 93%, about 94%, about 95%, about 96%, about 97%,

about 98%, about 99%, to about 100%, and any range therein, of the nucleobase sequence is capable of base-pairing with at least one single or double stranded nucleic acid molecule during hybridization. In certain embodiments, the term "substantially complementary" refers to at least one nucleic acid that may hybridize to at least one nucleic acid strand or duplex in stringent conditions. In certain embodiments, a "partly complementary" nucleic acid comprises at least one sequence that may hybridize in low stringency conditions to at least one single or double stranded nucleic acid, or contains at least one sequence in which less than about 70% of the nucleobase sequence is capable of base-pairing with at least one single or double stranded nucleic acid molecule during hybridization.

[00139] As used herein, "hybridization", "hybridizes" or "capable of hybridizing" is understood to mean the forming of a double or triple stranded molecule or a molecule with partial double or triple stranded nature. The term "hybridization", "hybridize(s)" or "capable of hybridizing" encompasses the terms "stringent condition(s)" or "high stringency" and the terms "low stringency" or "low stringency condition(s)".

[00140] As used herein "stringent condition(s)" or "high stringency" are those that allow hybridization between or within one or more nucleic acid strand(s) containing complementary sequence(s), but precludes hybridization of random sequences. Stringent conditions tolerate little, if any, mismatch between a nucleic acid and a target strand. Such conditions are well known to those of ordinary skill in the art, and are preferred for applications requiring high selectivity. Non-limiting applications include isolating at least one nucleic acid, such as a gene or nucleic acid segment thereof, or detecting at least one specific mRNA transcript or nucleic acid segment thereof, and the like.

[00141] Stringent conditions may comprise low salt and/or high temperature conditions, such as provided by about 0.02 M to about 0.15 M NaCl at temperatures of about 50°C to about 70°C. It is understood that the temperature and ionic strength of a desired stringency are determined in part by the length of the particular nucleic acid(s), the length and nucleobase content of the target sequence(s), the charge composition of the nucleic acid(s), and to the presence of formamide, tetramethylammonium chloride or other solvent(s) in the hybridization mixture. It is generally appreciated that conditions may be rendered more stringent, such as, for example, the addition of increasing amounts of formamide.

10041007-010702

[00142] It is also understood that these ranges, compositions and conditions for hybridization are mentioned by way of non-limiting example only, and that the desired stringency for a particular hybridization reaction is often determined empirically by comparison to one or more positive or negative controls. Depending on the application envisioned it is preferred to employ varying conditions of hybridization to achieve varying degrees of selectivity of the nucleic acid(s) towards target sequence(s). In a non-limiting example, identification or isolation of related target nucleic acid(s) that do not hybridize to a nucleic acid under stringent conditions may be achieved by hybridization at low temperature and/or high ionic strength. Such conditions are termed "low stringency" or "low stringency conditions", and non-limiting examples of low stringency include hybridization performed at about 0.15 M to about 0.9 M NaCl at a temperature range of about 20°C to about 50°C. Of course, it is within the skill of one in the art to further modify the low or high stringency conditions to suite a particular application.

[00143] One or more nucleic acid(s) may comprise, or be composed entirely of, at least one derivative or mimic of at least one nucleobase, a nucleobase linker moiety and/or backbone moiety that may be present in a naturally occurring nucleic acid. As used herein a "derivative" refers to a chemically modified or altered form of a naturally occurring molecule, while the terms "mimic" or "analog" refers to a molecule that may or may not structurally resemble a naturally occurring molecule, but functions similarly to the naturally occurring molecule. As used herein, a "moiety" generally refers to a smaller chemical or molecular component of a larger chemical or molecular structure, and is encompassed by the term "molecule."

[00144] As used herein a "nucleobase" refers to a naturally occurring heterocyclic base, such as A, T, G, C or U ("naturally occurring nucleobase(s)"), found in at least one naturally occurring nucleic acid (*i.e.* DNA and RNA), and their naturally or non-naturally occurring derivatives and mimics. Non-limiting examples of nucleobases include purines and pyrimidines, as well as derivatives and mimics thereof, which generally forms one or more hydrogen bonds ("anneal" or "hybridize") with at least one naturally occurring nucleobase in manner that may substitute for naturally occurring nucleobase pairing (*e.g.* the hydrogen bonding between A and T, G and C, and A and U).

[00145] Nucleobase, nucleoside and nucleotide mimics or derivatives are well known in the art, and have been described in exemplary references such as, for

example, Scheit, Nucleotide Analogs (John Wiley, New York, 1980), incorporated herein by reference. "Purine" and "pyrimidine" nucleobases encompass naturally occurring purine and pyrimidine nucleobases and also derivatives and mimics thereof, including but not limited to, those purines and pyrimidines substituted by one or more of alkyl, carboxyalkyl, amino, hydroxyl, halogen (*i.e.* fluoro, chloro, bromo, or iodo), thiol, or alkylthiol wherein the alkyl group comprises of from about 1, about 2, about 3, about 4, about 5, to about 6 carbon atoms. Non-limiting examples of purines and pyrimidines include deazapurines, 2,6-diaminopurine, 5-fluorouracil, xanthine, hypoxanthine, 8-bromoguanine, 8-chloroguanine, bromothymine, 8-aminoguanine, 8-hydroxyguanine, 8-methylguanine, 8-thioguanine, azaguanines, 2-aminopurine, 5-ethylcytosine, 5-methylcytosine, 5-bromouracil, 5-ethyluracil, 5-iodouracil, 5-chlorouracil, 5-propyluracil, thiouracil, 2-methyladenine, methylthioadenine, N,N-dimethyladenine, azaadenines, 8-bromoadenine, 8-hydroxyadenine, 6-hydroxyaminopurine, 6-thiopurine, 4-(6-aminoethyl)cytosine), and the like. Purine and pyrimidine derivatives and mimics are well known in the art.

[00146] As used herein, "nucleoside" refers to an individual chemical unit comprising a nucleobase covalently attached to a nucleobase linker moiety. A non-limiting example of a "nucleobase linker moiety" is a sugar comprising 5-carbon atoms (a "5-carbon sugar"), including but not limited to deoxyribose, ribose or arabinose, and derivatives or mimics of 5-carbon sugars. Non-limiting examples of derivatives or mimics of 5-carbon sugars include 2'-fluoro-2'-deoxyribose or carbocyclic sugars where a carbon is substituted for the oxygen atom in the sugar ring. By way of non-limiting example, nucleosides comprising purine (*i.e.* A and G) or 7-deazapurine nucleobases typically covalently attach the 9 position of the purine or 7-deazapurine to the 1'-position of a 5-carbon sugar. In another non-limiting example, nucleosides comprising pyrimidine nucleobases (*i.e.* C, T or U) typically covalently attach the 1 position of the pyrimidine to 1'-position of a 5-carbon sugar (Kornberg and Baker, DNA Replication, 2nd Ed. (Freeman, San Francisco, 1992). However, other types of covalent attachments of a nucleobase to a nucleobase linker moiety are known in the art, and non-limiting examples are described herein.

[00147] As used herein, a "nucleotide" refers to a nucleoside further comprising a "backbone moiety" generally used for the covalent attachment of one or



more nucleotides to another molecule or to each other to form one or more nucleic acids. The "backbone moiety" in naturally occurring nucleotides typically comprises a phosphorus moiety, which is covalently attached to a 5-carbon sugar. The attachment of the backbone moiety typically occurs at either the 3'- or 5'-position of the 5-carbon sugar. However, other types of attachments are known in the art, particularly when the nucleotide comprises derivatives or mimics of a naturally occurring 5-carbon sugar or phosphorus moiety, and non-limiting examples are described herein.

[00148] A non-limiting example of a nucleic acid comprising such nucleoside or nucleotide derivatives and mimics is a "polyether nucleic acid", described in U.S. Patent Serial No. 5,908,845, incorporated herein by reference, wherein one or more nucleobases are linked to chiral carbon atoms in a polyether backbone. Another example of a nucleic acid comprising nucleoside or nucleotide derivatives or mimics is a "peptide nucleic acid", also known as a "PNA", "peptide-based nucleic acid mimics" or "PENAMs", described in U.S. Patent Serial Nos. 5,786,461, 5,891,625, 5,773,571, 5,766,855, 5,736,336, 5,719,262, 5,714,331, 5,539,082, and WO 92/20702, each of which is incorporated herein by reference. A peptide nucleic acid generally comprises at least one nucleobase and at least one nucleobase linker moiety that is either not a 5-carbon sugar and/or at least one backbone moiety that is not a phosphate backbone moiety. Examples of nucleobase linker moieties described for PNAs include aza nitrogen atoms, amido and/or ureido tethers (see for example, U.S. Patent No. 5,539,082). Examples of backbone moieties described for PNAs include an aminoethylglycine, polyamide, polyethyl, polythioamide, polysulfonamide or polysulfonamide backbone moiety.

[00149] Peptide nucleic acids generally have enhanced sequence specificity, binding properties, and resistance to enzymatic degradation in comparison to molecules such as DNA and RNA (Egholm *et al.*, Nature 1993, 365, 566; PCT/EP/01219). In addition, U.S. Patent Nos. 5,766,855, 5,719,262, 5,714,331 and 5,736,336 describe PNAs comprising naturally and non-naturally occurring nucleobases and alkylamine side chains with further improvements in sequence specificity, solubility and binding affinity. These properties promote double or triple helix formation between a target nucleic acid and the PNA.

[00150] U.S. Patent No. 5,641,625 describes that the binding of a PNA may to a target sequence has applications the creation of PNA probes to nucleotide sequences,

modulating (*i.e.* enhancing or reducing) gene expression by binding of a PNA to an expressed nucleotide sequence, and cleavage of specific dsDNA molecules. In certain embodiments, nucleic acid analogues such as one or more peptide nucleic acids may be used to inhibit nucleic acid amplification, such as in PCR, to reduce false positives and discriminate between single base mutants, as described in U.S. Patent Serial No. 5,891,625.

[00151] U.S. Patent 5,786,461 describes PNAs with amino acid side chains attached to the PNA backbone to enhance solubility. The neutrality of the PNA backbone may contribute to the thermal stability of PNA/DNA and PNA/RNA duplexes by reducing charge repulsion. The melting temperature of PNA containing duplexes, or temperature at which the strands of the duplex release into single stranded molecules, has been described as less dependent upon salt concentration.

[00152] One method for increasing amount of cellular uptake property of PNAs is to attach a lipophilic group. U.S. application Ser. No. 117,363, filed Sep. 3, 1993, describes several alkylamino functionalities and their use in the attachment of such pendant groups to oligonucleosides. U.S. application Ser. No. 07/943,516, filed Sep. 11, 1992, and its corresponding published PCT application WO 94/06815, describe other novel amine-containing compounds and their incorporation into oligonucleotides for, *inter alia*, the purposes of enhancing cellular uptake, increasing lipophilicity, causing greater cellular retention and increasing the distribution of the compound within the cell.

[00153] Additional non-limiting examples of nucleosides, nucleotides or nucleic acids comprising 5-carbon sugar and/or backbone moiety derivatives or mimics are well known in the art.

[00154] In certain aspect, the present invention concerns at least one nucleic acid that is an isolated nucleic acid. As used herein, the term "isolated nucleic acid" refers to at least one nucleic acid molecule that has been isolated free of, or is otherwise free of, the bulk of the total genomic and transcribed nucleic acids of one or more cells, particularly plant cells, and more particularly *Ginkgo biloba* cells. In certain embodiments, "isolated nucleic acid" refers to a nucleic acid that has been isolated free of, or is otherwise free of, bulk of cellular components and macromolecules such as lipids, proteins, small biological molecules, and the like. As different species may have a RNA or a DNA containing genome, the term "isolated nucleic acid" encompasses both the terms

10041007.010702

"isolated DNA" and "isolated RNA". Thus, the isolated nucleic acid may comprise a RNA or DNA molecule isolated from, or otherwise free of, the bulk of total RNA, DNA or other nucleic acids of a particular species. As used herein, an isolated nucleic acid isolated from a particular species is referred to as a "species specific nucleic acid." When designating a nucleic acid isolated from a particular species, such as human, such a type of nucleic acid may be identified by the name of the species. For example, a nucleic acid isolated from one or more humans would be an "isolated human nucleic acid", a nucleic acid isolated from *Ginkgo biloba* would be an "isolated *Ginkgo biloba* nucleic acid", and the like.

**[00155]** Of course, more than one copy of an isolated nucleic acid may be isolated from biological material, or produced *in vitro*, using standard techniques that are known to those of skill in the art. In particular embodiments, the isolated nucleic acid is capable of expressing a protein, polypeptide or peptide that has diterpene synthase activity, such as levopimaradiene synthase activity. In other embodiments, the isolated nucleic acid comprises an isolated levopimaradiene synthase gene.

**[00156]** Herein certain embodiments, a "gene" refers to a nucleic acid that is transcribed. As used herein, a "gene segment" is a nucleic acid segment of a gene. In certain aspects, the gene includes regulatory sequences involved in transcription, or message production or composition. In particular embodiments, the gene comprises transcribed sequences that encode for a protein, polypeptide or peptide. In other particular aspects, the gene comprises a levopimaradiene synthase nucleic acid, and/or encodes a levopimaradiene synthase polypeptide or peptide coding sequences. In keeping with the terminology described herein, an "isolated gene" may comprise transcribed nucleic acid(s), regulatory sequences, coding sequences, or the like, isolated substantially away from other such sequences, such as other naturally occurring genes, regulatory sequences, polypeptide or peptide encoding sequences, *etc.* In this respect, the term "gene" is used for simplicity to refer to a nucleic acid comprising a nucleotide sequence that is transcribed, and the complement thereof. In particular aspects, the transcribed nucleotide sequence comprises at least one functional protein, polypeptide and/or peptide encoding unit. As is understood by those in the art, this function term "gene" includes both genomic sequences, RNA or cDNA sequences or smaller engineered nucleic acid segments, including nucleic acid segments of a non-transcribed part of a gene, including but not limited to the non-transcribed promoter or enhancer regions of a gene. Smaller engineered gene nucleic acid

segments may express, or may be adapted to express using nucleic acid manipulation technology, proteins, polypeptides, domains, peptides, fusion proteins, mutants and/or such like.

**[00157]** "Isolated substantially away from other coding sequences" means that the gene of interest, in this case the levopimaradiene synthase gene(s), forms the significant part of the coding region of the nucleic acid, or that the nucleic acid does not contain large portions of naturally-occurring coding nucleic acids, such as large chromosomal fragments, other functional genes, RNA or cDNA coding regions. Of course, this refers to the nucleic acid as originally isolated, and does not exclude genes or coding regions later added to the nucleic acid by the hand of man.

**[00158]** In certain embodiments, the nucleic acid is a nucleic acid segment. As used herein, the term "nucleic acid segment", are smaller fragments of a nucleic acid, such as for non-limiting example, those that encode only part of the levopimaradiene synthase peptide or polypeptide sequence. Thus, a "nucleic acid segment" may comprise any part of the levopimaradiene synthase gene sequence(s), of from about 2 nucleotides to the full length of the levopimaradiene synthase peptide or polypeptide encoding region. In certain embodiments, the "nucleic acid segment" encompasses the full length levopimaradiene synthase gene(s) sequence. In particular embodiments, the nucleic acid comprises any part of the SEQ.ID.NO:1 sequence(s), of from about 2 nucleotides to the full length of the sequence disclosed in SEQ.ID.NO:1.

**[00159]** The nucleic acid(s) of the present invention, regardless of the length of the sequence itself, may be combined with other nucleic acid sequences, including but not limited to, promoters, enhancers, polyadenylation signals, restriction enzyme sites, multiple cloning sites, coding segments, and the like, to create one or more nucleic acid construct(s). The length overall length may vary considerably between nucleic acid constructs. Thus, a nucleic acid segment of almost any length may be employed, with the total length preferably being limited by the ease of preparation or use in the intended recombinant nucleic acid protocol.

**[00160]** In a non-limiting example, one or more nucleic acid constructs may be prepared that include a contiguous stretch of nucleotides identical to or complementary to SEQ.ID.NO:1. A nucleic acid construct may be about 3, about 5, about 8, about 10 to about 14, or about 15, about 20, about 30, about 40, about 50, about 100, about 200, about

500, about 1,000, about 2,000, about 3,000, about 5,000, about 10,000, about 15,000, about 20,000, about 30,000, about 50,000, about 100,000, about 250,000, about 500,000, about 750,000, to about 1,000,000 nucleotides in length, as well as constructs of greater size, up to and including chromosomal sizes (including all intermediate lengths and intermediate ranges), given the advent of nucleic acids constructs such as a yeast artificial chromosome are known to those of ordinary skill in the art. It is readily understood that "intermediate lengths" and "intermediate ranges", as used herein, means any length or range including or between the quoted values (*i.e.* all integers including and between such values). Non-limiting examples of intermediate lengths include about 11, about 12, about 13, about 16, about 17, about 18, about 19, *etc.*; about 21, about 22, about 23, *etc.*; about 31, about 32, *etc.*; about 51, about 52, about 53, *etc.*; about 101, about 102, about 103, *etc.*; about 151, about 152, about 153, *etc.*; about 1,001, about 1002, *etc.*; about 50,001, about 50,002, *etc.*; about 750,001, about 750,002, *etc.*; about 1,000,001, about 1,000,002, *etc.* Non-limiting examples of intermediate ranges include about 3 to about 32, about 150 to about 500,001, about 3,032 to about 7,145, about 5,000 to about 15,000, about 20,007 to about 1,000,003, *etc.*

[00161] In particular embodiments, the invention concerns one or more recombinant vector(s) comprising nucleic acid sequences that encode a levopimaradiene synthase protein, polypeptide or peptide that includes within its amino acid sequence a contiguous amino acid sequence in accordance with, or essentially as set forth in, SEQ.ID.NO:2, corresponding to *Ginkgo biloba* levopimaradiene synthase. In particular aspects, the recombinant vectors are DNA vectors.

[00162] The term "a sequence essentially as set forth in SEQ.ID.NO:2" means that the sequence substantially corresponds to a portion of SEQ.ID.NO:2 and has relatively few amino acids that are not identical to, or a biologically functional equivalent of, the amino acids of SEQ.ID.NO:2. Accordingly, a sequence that has between about 70% and about 80%; or more preferably, between about 81% and about 90%; or even more preferably, between about 91% and about 99%; of amino acids that are identical or functionally equivalent to the amino acids of SEQ.ID.NO:2 is a sequence that is "essentially as set forth in SEQ.ID.NO:2" Thus, "a sequence essentially as set forth in SEQ.ID.NO:1" encompasses nucleic acids, nucleic acid segments, and genes that comprise part or all of the nucleic acid sequences as set forth in SEQ.ID.NO:1, wherein the

sequence that has between about 70% and about 80%; or more preferably, between about 81% and about 90%; or even more preferably, between about 91% and about 99%; of amino acids that are identical or functionally equivalent to the nucleic acids of SEQ.ID.NO:1.

**[00163]** The term "biologically functional equivalent" is well understood in the art and is further defined in detail herein. A nucleic acid sequence encoding a polypeptide that performs an equivalent function to the polypeptide of amino acid SEQ.ID.NO:2 is a sequence that is a "biologically functional equivalent" protein, polypeptide or peptide. Likewise, the nucleic acid sequence encoding the biologically functional equivalent polypeptide is also contemplated within the scope of the invention.

**[00164]** The term "conservatively modified variants" applies to both amino acid and nucleic acid sequences. With respect to particular nucleic acid sequences, conservatively modified variants refers to those nucleic acids which encode identical or conservatively modified variants of the amino acid sequences. Because of the degeneracy of the genetic code, a large number of functionally identical nucleic acids encode any given protein. For instance, the codons GCA, GCC, GCG and GCU all encode the amino acid alanine. Thus, at every position where an alanine is specified by a codon, the codon is altered to any of the corresponding codons described without altering the encoded polypeptide. Such nucleic acid variations are "silent variations" and represent one species of conservatively modified variation. Every nucleic acid sequence herein which encodes a polypeptide also describes every possible silent variation of the nucleic acid. One of ordinary skill recognizes that each codon in a nucleic acid (except AUG, which is ordinarily the only codon for methionine; and UGG, which is ordinarily the only codon for tryptophan) is modified to yield a functionally identical molecule. Accordingly, each silent variation of a nucleic acid which encodes a polypeptide of the present invention is implicit in each described polypeptide sequence and incorporated herein by reference.

**[00165]** As to amino acid sequences, one of skill recognizes that individual substitutions, deletions or additions to a nucleic acid, peptide, polypeptide, or protein sequence which alters, adds or deletes a single amino acid or a small percentage of amino acids in the encoded sequence is a "conservatively modified variant" where the alteration results in the substitution of an amino acid with a chemically similar amino acid. Thus, any number of amino acid residues selected from the group of integers consisting of from

1 to 15 is so altered. Thus, for example, 1, 2, 3, 4, 5, 7, or 10 alterations are made. Conservatively modified variants typically provide similar biological activity as the unmodified polypeptide sequence from which they are derived. For example, substrate specificity, enzyme activity, or ligand/receptor binding is generally at least 30%, 40%, 50%, 60%, 70%, 80%, or 90% of the native protein for its native substrate. Conservative substitution tables providing functionally similar amino acids are well known in the art.

[00166] In certain other embodiments, the invention concerns at least one recombinant vector that include within its sequence a nucleic acid sequence essentially as set forth in SEQ.ID.NO:1. In particular embodiments, the recombinant vector comprises DNA sequences that encode protein(s), polypeptide(s) or peptide(s) exhibiting levopimaradiene synthase activity.

[00167] It also understood that amino acid sequences or nucleic acid sequences may include additional residues, such as additional N- or C-terminal amino acids or 5' or 3' sequences, or various combinations thereof, and yet still be essentially as set forth in one of the sequences disclosed herein, so long as the sequence meets the criteria set forth above, including the maintenance of biological protein, polypeptide or peptide activity where expression of a proteinaceous composition is concerned. The addition of terminal sequences particularly applies to nucleic acid sequences that may, for example, include various non-coding sequences flanking either of the 5' and/or 3' portions of the coding region or may include various internal sequences, *i.e.*, introns, which are known to occur within genes.

[00168] Excepting intronic and flanking regions, and allowing for the degeneracy of the genetic code, nucleic acid sequences that have between about 70% and about 79%; or more preferably, between about 80% and about 89%; or even more particularly, between about 90% and about 99%; of nucleotides that are identical to the nucleotides of SEQ.ID.NO:1 are nucleic acid sequences that are "essentially as set forth in SEQ.ID.NO:1".

[00169] It also understood that this invention is not limited to the particular nucleic acid of SEQ.ID.NO:1 or amino acid sequences of SEQ.ID.NO:2. Recombinant vectors and isolated nucleic acid segments may therefore variously include these coding regions themselves, coding regions bearing selected alterations or modifications in the basic coding region, and they may encode larger polypeptides or peptides that nevertheless

include such coding regions or may encode biologically functional equivalent proteins, polypeptide or peptides that have variant amino acids sequences.

**[00170]** The nucleic acids of the present invention encompass biologically functional equivalent levopimaradiene synthase proteins, polypeptides, or peptides. Such sequences may arise as a consequence of codon redundancy or functional equivalency that are known to occur naturally within nucleic acid sequences or the proteins, polypeptides or peptides thus encoded. Alternatively, functionally equivalent proteins, polypeptides or peptides may be created *via* the application of recombinant DNA technology, in which changes in the protein, polypeptide or peptide structure may be engineered, based on considerations of the properties of the amino acids being exchanged. Changes designed by man may be introduced, for example, through the application of site-directed mutagenesis techniques as discussed herein below, *e.g.*, to introduce improvements or alterations to the antigenicity of the protein, polypeptide or peptide, or to test mutants in order to examine levopimaradiene synthase protein, polypeptide or peptide activity at the molecular level.

**[00171]** Fusion proteins, polypeptides or peptides may be prepared, *e.g.*, where the levopimaradiene synthase-coding regions are aligned within the same expression unit with other proteins, polypeptides or peptides having desired functions. Non-limiting examples of such desired functions of expression sequences include purification or immunodetection purposes for the added expression sequences, *e.g.*, proteinaceous compositions that may be purified by affinity chromatography or the enzyme labeling of coding regions, respectively.

**[00172]** Encompassed by the invention are nucleic acid sequences encoding relatively small peptides or fusion peptides, such as, for example, peptides of from about 3, about 4, about 5, about 6, about 7, about 8, about 9, about 10, about 11, about 12, about 13, about 14, about 15, about 16, about 17, about 18, about 19, about 20, about 21, about 22, about 23, about 24, about 25, about 26, about 27, about 28, about 29, about 30, about 31, about 32, about 33, about 34, about 35, about 35, about 36, about 37, about 38, about 39, about 40, about 41, about 42, about 43, about 44, about 45, about 46, about 47, about 48, about 49, about 50, about 51, about 52, about 53, about 54, about 55, about 56, about 57, about 58, about 59, about 60, about 61, about 62, about 63, about 64, about 65, about 66, about 67, about 68, about 69, about 70, about 71, about 72, about 73, about 74, about



75, about 76, about 77, about 78, about 79, about 80, about 81, about 82, about 83, about 84, about 85, about 86, about 87, about 88, about 89, about 90, about 91, about 92, about 93, about 94, about 95, about 96, about 97, about 98, about 99, to about 100 amino acids in length, or more preferably, of from about 15 to about 30 amino acids in length; as set forth in SEQ.ID.NO:2 and also larger polypeptides up to and including proteins corresponding to the full-length sequences set forth in SEQ.ID.NO:2.

[00173] As used herein an "organism" may be a prokaryote, eukaryote, virus and the like. As used herein the term "sequence" encompasses both the terms "nucleic acid" and "proteinaceous" or "proteinaceous composition." As used herein, the term "proteinaceous composition" encompasses the terms "protein", "polypeptide" and "peptide." As used herein "artificial sequence" refers to a sequence of a nucleic acid not derived from sequence naturally occurring at a genetic locus, as well as the sequence of any proteins, polypeptides or peptides encoded by such a nucleic acid. A "synthetic sequence", refers to a nucleic acid or proteinaceous composition produced by chemical synthesis *in vitro*, rather than enzymatic production *in vitro* (i.e. an "enzymatically produced" sequence) or biological production *in vivo* (i.e. a "biologically produced" sequence).

### VIII. Methods for Plant Transformation

[00174] Suitable methods for plant transformation for use with the current invention are believed to include virtually any method by which DNA is introduced into a cell, such as by direct delivery of DNA such as by PEG-mediated transformation of protoplasts (Omirulleh *et al.*, 1993), by desiccation/inhibition-mediated DNA uptake (Potrykus *et al.*, 1985), by electroporation (U.S. Patent No. 5,384,253, specifically incorporated herein by reference in its entirety), by agitation with silicon carbide fibers (Kaepler *et al.*, 1990; U.S. Patent No. 5,302,523, specifically incorporated herein by reference in its entirety; and U.S. Patent No. 5,464,765, specifically incorporated herein by reference in its entirety), by *Agrobacterium*-mediated transformation (U.S. Patent No. 5,591,616 and U.S. Patent No. 5,563,055; both specifically incorporated herein by reference) and by acceleration of DNA coated particles (U.S. Patent No. 5,550,318; U.S. Patent No. 5,538,877; and U.S. Patent No. 5,538,880; each specifically incorporated herein by reference in its entirety), and the like. Through the application of techniques

such as these, maize cells as well as those of virtually any other plant species may be stably transformed, and these cells developed into transgenic plants. In certain embodiments, acceleration methods are preferred and include, for example, microprojectile bombardment and the like.

[00175] A transgenic plant may require seed propagation, and in such instances, a seed of the transgenic plant embodies the recombinant gene therein. In the case of Ginkgo, the genetic content of the seeds is particularly enriched for ginkgolide production. Thus, the seed of a transgenic plant is characterized by increased amounts of a ginkgolide and is a reasonable means to propagate the transgenic plant that is the resource for the sought-after ginkgolide.

#### A. Electroporation

[00176] Where one wishes to introduce DNA by means of electroporation, it is contemplated that the method of Krzyzek *et al.* (U.S. Patent No. 5,384,253, incorporated herein by reference in its entirety) is particularly advantageous. In this method, certain cell wall-degrading enzymes, such as pectin-degrading enzymes, are employed to render the target recipient cells more susceptible to transformation by electroporation than untreated cells. Alternatively, recipient cells are made more susceptible to transformation by mechanical wounding.

[00177] To effect transformation by electroporation, one may employ either friable tissues, such as a suspension culture of cells or embryogenic callus or alternatively one may transform immature embryos or other organized tissue directly. In this technique, one would partially degrade the cell walls of the chosen cells by exposing them to pectin-degrading enzymes (pectolyases) or mechanically wounding in a controlled manner. Examples of some species which have been transformed by electroporation of intact cells include maize (U.S. Patent No. 5,384,253; Rhodes *et al.*, 1995; D'Halluin *et al.*, 1992), wheat (Zhou *et al.*, 1993), tomato (Hou and Lin, 1996), soybean (Christou *et al.*, 1987) and tobacco (Lee *et al.*, 1989).

[00178] One also may employ protoplasts for electroporation transformation of plants (Bates, 1994; Lazzeri, 1995). For example, the generation of transgenic soybean plants by electroporation of cotyledon-derived protoplasts is described by Dhir and Widholm in Intl. Patent Appl. Publ. No. WO 9217598 (specifically incorporated herein by

reference). Other examples of species for which protoplast transformation has been described include barley (Lazerri, 1995), sorghum (Battraw et al., 1991), maize (Bhattacharjee et al., 1997), wheat (He et al., 1994) and tomato (Tsukada, 1989).

## **B. Microprojectile Bombardment**

[00179] A preferred method for delivering transforming DNA segments to plant cells in accordance with the invention is microprojectile bombardment (U.S. Patent No. 5,550,318; U.S. Patent No. 5,538,880; U.S. Patent No. 5,610,042; and PCT Application WO 94/09699; each of which is specifically incorporated herein by reference in its entirety). In this method, particles may be coated with nucleic acids and delivered into cells by a propelling force. Exemplary particles include those comprised of tungsten, platinum, and preferably, gold. It is contemplated that in some instances DNA precipitation onto metal particles would not be necessary for DNA delivery to a recipient cell using microprojectile bombardment. However, it is contemplated that particles may contain DNA rather than be coated with DNA. Hence, it is proposed that DNA-coated particles may increase the level of DNA delivery *via* particle bombardment but are not, in and of themselves, necessary.

[00180] For the bombardment, cells in suspension are concentrated on filters or solid culture medium. Alternatively, immature embryos or other target cells may be arranged on solid culture medium. The cells to be bombarded are positioned at an appropriate distance below the macroprojectile stopping plate.

[00181] An illustrative embodiment of a method for delivering DNA into plant cells by acceleration is the Biolistics Particle Delivery System, which is used to propel particles coated with DNA or cells through a screen, such as a stainless steel or Nytex screen, onto a filter surface covered with monocot plant cells cultured in suspension. The screen disperses the particles so that they are not delivered to the recipient cells in large aggregates. It is believed that a screen intervening between the projectile apparatus and the cells to be bombarded reduces the size of projectiles aggregate and may contribute to a higher frequency of transformation by reducing the damage inflicted on the recipient cells by projectiles that are too large.

[00182] Microprojectile bombardment techniques are widely applicable, and may be used to transform virtually any plant species. Examples of species for which have

been transformed by microprojectile bombardment include monocot species such as maize (PCT Application WO 95/06128), barley (Ritala et al., 1994; Hensgens et al., 1993), wheat (U.S. Patent No. 5,563,055, specifically incorporated herein by reference in its entirety), rice (Hensgens et al., 1993), oat (Torbet et al., 1995; Torbet et al., 1998), rye (Hensgens et al., 1993), sugarcane (Bower et al., 1992), and sorghum (Casas et al., 1993; Hagio et al., 1991); as well as a number of dicots including tobacco (Tomes et al., 1990; Buisning and Benbow, 1994), soybean (U.S. Patent No. 5,322,783, specifically incorporated herein by reference in its entirety), sunflower (Knittel et al. 1994), peanut (Singsit et al., 1997), cotton (McCabe and Martinell, 1993), tomato (VanEck et al. 1995), and legumes in general (U.S. Patent No. 5,563,055, specifically incorporated herein by reference in its entirety).

### **C. *Agrobacterium*-mediated Transformation**

[00183] *Agrobacterium*-mediated transfer is a widely applicable system for introducing genes into plant cells because the DNA is introduced into whole plant tissues, thereby bypassing the need for regeneration of an intact plant from a protoplast. The use of *Agrobacterium*-mediated plant integrating vectors to introduce DNA into plant cells is well known in the art. See, for example, the methods described by Fraley *et al.*, (1985), Rogers *et al.*, (1987) and U.S. Patent No. 5,563,055, specifically incorporated herein by reference in its entirety.

[00184] *Agrobacterium*-mediated transformation is most efficient in dicotyledonous plants and is the preferable method for transformation of dicots, including *Arabidopsis*, tobacco, tomato, and potato. Indeed, while *Agrobacterium*-mediated transformation has been routinely used with dicotyledonous plants for a number of years, it has only recently become applicable to monocotyledonous plants. Advances in *Agrobacterium*-mediated transformation techniques have now made the technique applicable to nearly all monocotyledonous plants. For example, *Agrobacterium*-mediated transformation techniques have now been applied to rice (Hiei *et al.*, 1997; Zhang *et al.*, 1997; U.S. Patent No. 5,591,616, specifically incorporated herein by reference in its entirety), wheat (McCormac *et al.*, 1998), barley (Tingay *et al.*, 1997; McCormac *et al.*, 1998), and maize (Ishidia *et al.*, 1996).

[00185] Modern *Agrobacterium* transformation vectors are capable of replication in *E. coli* as well as *Agrobacterium*, allowing for convenient manipulations as

described (Klee *et al.*, 1985). Moreover, recent technological advances in vectors for *Agrobacterium*-mediated gene transfer have improved the arrangement of genes and restriction sites in the vectors to facilitate the construction of vectors capable of expressing various polypeptide coding genes. The vectors described (Rogers *et al.*, 1987) have convenient multi-linker regions flanked by a promoter and a polyadenylation site for direct expression of inserted polypeptide coding genes and are suitable for present purposes. In addition, *Agrobacterium* containing both armed and disarmed T<sub>i</sub> genes are used for the transformations. In those plant strains where *Agrobacterium*-mediated transformation is efficient, it is the method of choice because of the facile and defined nature of the gene transfer.

#### **D. Other Transformation Methods**

[00186] Transformation of plant protoplasts is achieved using methods based on calcium phosphate precipitation, polyethylene glycol treatment, electroporation, and combinations of these treatments (see, e.g., Potrykus *et al.*, 1985; Lorz *et al.*, 1985; Omirulleh *et al.*, 1993; Fromm *et al.*, 1986; Uchimiyu *et al.*, 1986; Callis *et al.*, 1987; Marcotte *et al.*, 1988).

[00187] Application of these systems to different plant strains depends upon the ability to regenerate that particular plant strain from protoplasts. Illustrative methods for the regeneration of cereals from protoplasts have been described (Fujimara *et al.*, 1985; Toriyama *et al.*, 1986; Yamada *et al.*, 1986; Abdullah *et al.*, 1986; Omirulleh *et al.*, 1993 and U.S. Patent No. 5,508,184; each specifically incorporated herein by reference in its entirety). Examples of the use of direct uptake transformation of cereal protoplasts include transformation of rice (Ghosh-Biswas *et al.*, 1994), sorghum (Battraw and Hall, 1991), barley (Lazerri, 1995), oat (Zheng and Edwards, 1990) and maize (Omirulleh *et al.*, 1993).

[00188] To transform plant strains that are not successfully regenerated from protoplasts, other ways to introduce DNA into intact cells or tissues are utilized. For example, regeneration of cereals from immature embryos or explants are effected as described (Vasil, 1989). Also, silicon carbide fiber-mediated transformation is used with or without protoplasting (Kaeppler, 1990; Kaeppler *et al.*, 1992; U.S. Patent No. 5,563,055, specifically incorporated herein by reference in its entirety). Transformation with this technique is accomplished by agitating silicon carbide fibers together with cells

in a DNA solution. DNA passively enters as the cell are punctured. This technique has been used successfully with, for example, the monocot cereals maize (PCT Application WO 95/06128, specifically incorporated herein by reference in its entirety; Thompson, 1995) and rice (Nagatani, 1997).

[00189] An embodiment of the present invention is to produce significant amounts of ginkgolide precursors and/or ginkgolide *in vivo* in *Ginkgo* or microorganisms such as *Saccharomyces cerevisiae*, *Escherichia coli*, *Candida albicans*, and the like. Cell suspension cultures of *Ginkgo biloba* are known in the art (Balz *et al.*, 1999; Fiehe *et al.*, 2000).

[00190] In a preferred embodiment, ginkgolide precursors and/or ginkgolides are produced *in vivo* by expressing a nucleic acid sequence which encodes *Ginkgo biloba* levopimaradiene synthase, which is a rate-limiting step in the ginkgolide biosynthesis. In another preferred embodiment, the expression is upregulated, or "overexpressed" compared to native levels in wild type. A skilled artisan is aware how to achieve overexpression, such as by controlling regulation of the *Ginkgo biloba* levopimaradiene synthase with a strong promoter, examples of which are known in the art. In another preferred embodiment, the promoter is an inducible promoter, such as GAL1.

## EXAMPLES

[00191] The following examples are included to demonstrate preferred embodiments of the invention. It should be appreciated by those skilled in the art that the techniques disclosed in the examples which follow represent techniques discovered by the inventor to function well in the practice of the invention, and thus is considered to constitute preferred modes for its practice. However, those of skill in the art should, in light of the present disclosure, appreciate that many changes made in the specific embodiments which are disclosed and maintain a like or similar result without departing from the concept, spirit and scope of the invention. More specifically, it is apparent that certain agents that are both chemically and physiologically related may be substituted for the agents described herein while the same or similar results would be achieved. All such similar substitutes and modifications apparent to those skilled in the art are deemed to be within the spirit, scope and concept of the invention as defined by the appended claims.

## EXAMPLE 1

### METHODS-PLANT MATERIALS, SUBSTRATES, AND REAGENTS

[00192] *Ginkgo biloba* "white nut" seeds were purchased from Dynasty Supermarket (Houston, Texas). The seeds were stored at 4 °C for several days before sowing. Embryos were cultivated under aseptic conditions in an agar medium supplemented with D-glucose, L-glutamine, and Heller's salts at room temperature in the dark for four to six weeks (Schwarz, 1994). Synthesis of geranylgeraniol was performed as indicated in Ruan (1999) and Coates *et al.* (1978). Synthesis of geranylgeranyl diphosphate was performed as indicated by Corey *et al.* (1972), Davisson *et al.* (1985), and Davisson *et al.* (1986). Levopimarol was synthesized from levopimaric acid (Helix Biotech; New Westminster, British Columbia, Canada) according to procedures of Abad *et al.* (1985), Gigante *et al.* (1999), and Ayer and Talamas (1988). (6*E*,10*E*)-Geranyllinalool and pyridinium dichlorochromate were obtained from Fluka. All other reagents were obtained from either Sigma/Aldrich or Fisher Scientific, unless otherwise noted. Dichloromethane, dimethylformamide, methansulfonyl chloride, triethylamine, and toluene were freshly distilled over calcium hydride; tetrahydrofuran was freshly distilled over Na/benzophenone. Ammonium molybdate, ascorbic acid, tetrabutylammonium hydroxide, and citric acid were from ACROS. HP 20 polyaromatic dianion resin (250-850 µm) was purchased from Supelco.

## EXAMPLE 2

### METHODS- *G. BILOBA* mRNA ISOLATION, cDNA LIBRARY

#### CONSTRUCTION, AND QUALITY ASSESSMENT

[00193] At active growth, embryonic roots were harvested at 33 days (HSG1) and 40 days (HSG2) and snap-frozen in liquid nitrogen. Total RNA was isolated using a modified protocol for total RNA isolation from pine trees (Chang *et al.*, 1993) in which an acidic phenol/chloroform (1:1) extraction was included prior to ethanol precipitation. Poly(A)<sup>+</sup> RNA was selected with oligo(dT)cellulose (Life Technologies MessageMaker Poly(A)<sup>+</sup> Syringe Kit) and purified using Sephadex gel chromatography (Boehringer Mannheim Mini Quick Spin RNA Column) according to manufacturer's instructions. cDNA libraries were prepared using the SuperScript<sup>TM</sup> Plasmid System for cDNA Synthesis and Plasmid Cloning (Life Technologies). cDNA constructs (*Sall/Not I*)

1004-1007 010702

were subcloned into both the *E. coli* expression vector pSPORT1 and the centromeric yeast shuttle vector pRS316GAL (Liu and Krizek, 1992). The resultant plasmids were transformed by electroporation into ElectroMax™ DH10B Cells (Life Technologies). The number of transformants in each library varied from  $4.8 \times 10^5$  to  $3.2 \times 10^6$  with approximate insert size ranging from 200 to 2600 bp.

[00194] Complementation experiments were conducted to determine the quality of the libraries. *G. biloba* cDNA library HSG2 (10 µg) in pRS316GAL was transformed into the auxotrophic *S. cerevisiae* strain JBY575 (*MATa ura3-52-trp1-Δ63 leu2-3,112 his3-Δ200 ade2 Gal<sup>+</sup>*) (Corey *et al.*, 1996) using the lithium acetate method (Ito *et al.*, 1983), plated onto synthetic complete medium lacking uracil and supplemented with 2% glucose and 1.5% agar, and grown at 30 °C. A total of  $1.6 \times 10^5$  colonies were screened. Prototrophic clones were selected for growth by replica plating onto synthetic complete medium lacking leucine, tryptophan, or histidine and supplemented with 2% galactose and 1.5% agar, and incubated at 30 °C. The frequency of complementing cDNA was 1 in every 17,778 to 40,000 for *LEU2*, 1 in every 20,000 to 32,000 for *TRP1*, and 1 in every 22,800 to 32,000 for *HIS3*.

### EXAMPLE 3

#### METHODS- LEVOPIMARADIENE SYNTHASE GENE CLONING

[00195] PCR degenerate primers were designed according to sequence similarity between gymnosperm terpene synthases *Abies grandis* abietadiene synthase (SEQ.ID.NO:3), *Abies grandis* E-α-bisabolene synthase (GenBank Accession No. AF006195; SEQ.ID.NO:13 or GenBank Accession No. AF006194; SEQ.ID.NO:14; corresponding to the amino acid sequence in GenBank Accession No. AAC24192.1; SEQ.ID.NO:15 or AAC24191.1; SEQ.ID.NO:16, respectively), *Abies grandis* δ-selinene synthase (GenBank Accession No. U92266; SEQ.ID.NO:17; corresponding to the amino acid sequence in GenBank Accession No. AAC05727.1; SEQ.ID.NO:18); *Abies grandis* γ-humulene synthase (GenBank Accession No. U92267; SEQ.ID.NO:19; corresponding to the amino acid sequence in GenBank Accession No. AAC05728.1; SEQ.ID.NO:20), *Abies grandis* pinene synthase (GenBank Accession No. U87909; SEQ.ID.NO:21; corresponding to the amino acid sequence in GenBank Accession No. AAB71085.1; SEQ.ID.NO:22); *Abies grandis* (-)-4S-limonene synthase (GenBank Accession No.



AF006193; SEQ.ID.NO:23; corresponding to the amino acid sequence in GenBank Accession No. AAB70907.1; SEQ.ID.NO:24); *Abies grandis* myrcene synthase (GenBank Accession No. U87908; SEQ.ID.NO:25; corresponding to the amino acid sequence in GenBank Accession No. AAB71084.1; SEQ.ID.NO:26); *Abies grandis* (-)-limonene/(-)-alpha-pinene synthase (*agc11*) (GenBank Accession No. AF139207; SEQ.ID.NO:27; corresponding to the amino acid sequence in GenBank Accession No. AAF61455.1; SEQ.ID.NO:28); and *Taxus brevifolia* taxadiene synthase (SEQ.ID.NO:41).

**[00196]** PCR reactions were conducted on 50  $\mu$ L scale containing 200 ng cDNA, 5.0  $\mu$ L 10X PC2 buffer (500 mM Tris-HCl, pH=9.1, 160 mM (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>, 35 mM MgCl<sub>2</sub>), 4.0  $\mu$ L 2.5 mM dNTPs, and 5.0  $\mu$ L (20 pmol/ $\mu$ L) forward and reverse degenerate primers. The program employed a 4 min 95°C hot-start after which 0.5  $\mu$ L *Taq* DNA Polymerase (5.0 Units/ $\mu$ L, Fisher Biotech) was added to the PCR reaction, followed by 40 cycles with 1 min annealing using a temperature gradient from 68°C to 48°C (-0.5 °C/cycle), 3 min extension at 72°C, and 45 second denaturation at 95°C. The program was terminated with a 5 min extension at 72°C. An aliquot of each reaction (5  $\mu$ L) was analyzed on 2% agarose gel. The first round of PCR reactions employed the degenerate forward primer HSG1FP (5'-GCNTAYGAYACNGCNTGGGT-3'; SEQ.ID.NO:29). Combination with HSG6RP (5'-GCYTKRTANGTYTTNGTRTC-3'; SEQ.ID.NO:30) resulted in a 1907 bp fragment (HSG97), which was re-amplified, gel purified (QIAGEN), quantitated, and ligated into pGEM-T vector (50 ng/ $\mu$ L, Promega).

**[00197]** Conventional abbreviations are used in the primer sequences, wherein N is any base, K is G or T, Y is a pyrimidine, and R is a purine. The remainder of the sequence was obtained with specific primers HSG97.3FP (5'-ATGTGGTGGACTGGCAAGAG-3'; SEQ.ID.NO:5) and HSG97.3RP (5'-TAAAGATCGTCCAGAATAAC-3'; SEQ.ID.NO:6). A 1372 bp segment was excised with *Dra*I and *Bsr*G I. The DNA fragment (25 ng) was radiolabeled with  $\alpha$ -<sup>32</sup>P-dCTP using random oligonucleotide primers to probe  $3.0 \times 10^5$  colonies (cDNA library HSG2E) by colony hybridization (Ausubel *et al.*, 1999).

**[00198]** A total of 10 colonies were obtained, for which an additional round of screening yielded 47 hybridizing colonies. Six colonies were investigated further and restriction enzyme mapping indicated that three clones corresponded to the size of the expected full-length cDNA. Sequence analysis with forward primer T7 and reverse primer

SP6 indicated that these genes were putative diterpene cyclases based on homology to *Abies grandis* abietadiene synthase. All three clones exhibited identical 5' and 3' ends, therefore, one was selected for gene characterization.

[00199] Primers HSG100.1FP (5'-AACTGCCAGATGGCTCGTGG-3'; SEQ.ID.NO:7) and HSG100.2FP (5'-GGTGGAGTATGCTATAAAGT-3'; SEQ.ID.NO:8) were used along with HSG97.3FP to obtain the remaining sequence. Sequence data revealed that a 2681 bp cDNA (HSG100/pSPORT1) had been cloned, however, the initiation codon was absent. RNA ligase mediated rapid amplification of cDNA ends (FirstChoice™ RLM-RACE Kit, Ambion) was employed with outer gene specific primer HSG150OGS (5'-CAGAGCCGTC AATTGACGGAATTC-3'; SEQ.ID.NO:9) and inner gene specific primer HSG150IGS (5'-CATCGACGCTTGATTTCGATGTCG-3'; SEQ.ID.NO:10) to obtain the N-terminal sequence. The full-length clone (2705 bp) encoded an 873 amino acid open reading frame of 2622 bp with a predicted molecular weight of 100,289. Sequence alignment using the Clustal method indicated a 60% identity to *Abies grandis* abietadiene synthase, 46% to *Abies grandis* bisabolene synthase, and 41% to *Taxus brevifolia* taxadiene synthase (FIG. 4).

[00200] Sequence alignment in FIG. 4 was prepared with MegAlign (DNASTar, Madison, WI) using the Clustal method. Amino acid residues identical in at least three of the four synthases are shaded; hyphens indicate gaps inserted to maximize sequence alignment. Lines indicate aspartate-rich motifs, arrows designate regions targeted by degenerate PCR primers, and arrowheads identify N-terminal cleavage sites.

[00201] GbLS is *Ginkgo biloba* levopimaradiene synthase; AgAs is *Abies grandis* abietadiene synthase; AgBS is *Abies grandis* bioabolenne synthase; and TbTS is *Taxus brevifolia* taxadiene synthase.

## EXAMPLE 4

### METHODS- cDNA EXPRESSION AND ENZYMATIC ASSAY

[00202] Site-directed mutagenesis of HSG100/pSPORT1 with primers Ala<sup>2</sup> (5'-TTGCAAAGAGCACCCAGCCATTTT TTTTTCGACACCCGGGAATT CCGGACCGGT-3'; SEQ.ID.NO:11), Ser<sup>61</sup> (5'-TGGACGAGTCTCTGCAGCTGAC ATTTT TTTTTCGACCAATTCATCTCAGCCTT-3'; SEQ.ID.NO:12), Leu<sup>80</sup> (5'-

10041007.010702

20200702

TGATAATCCGCATTAAGCATTTTGTGTCGACTCCTCTGTGGAAGCTGAT-3'; SEQ.ID.NO:31), and Phe<sup>129</sup> (5'-TCGCCCATGGACTGAAACATT TTTTGTGTCGACTTCACCAATGTCTGGATTCT-3'; SEQ.ID.NO:40) was employed to incorporate a *SaI*I restriction site and a methionine initiation codon immediately upstream of Ala<sup>2</sup>, Ser<sup>61</sup>, Leu<sup>80</sup>, and Phe<sup>129</sup>. The plastid targeting sequence (e.g., N-terminal sequence) was removed by sequential digest with *SaI*I followed by *Not*I.

[00203] In specific embodiments, the Ala<sup>2</sup> mutant amino acid sequence (SEQ.ID.NO:33) is encoded by the nucleic acid sequence SEQ.ID.NO:32; the Ser<sup>61</sup> mutant amino acid sequence (SEQ.ID.NO:35) is encoded by the nucleic acid sequence of SEQ.ID.NO:34; the Leu<sup>80</sup> mutant amino acid sequence (SEQ.ID.NO:37) is encoded by the nucleic acid sequence of SEQ.ID.NO:36; and the Phe<sup>129</sup> mutant amino acid sequence (SEQ.ID.NO:39) is encoded by the nucleic acid sequence of SEQ.ID.NO:38. In a specific embodiment, an N-terminal truncation at any point in the amino acid sequence up to and including amino acid 129. In specific embodiments, alternative truncations are generated at the following sites: Cys<sup>55</sup>, Glu<sup>74</sup>, Glu<sup>76</sup>, or Val<sup>88</sup>, wherein the truncation site occurs just prior to the indicated amino acid (for example, between Asn<sup>54</sup> and Cys<sup>55</sup>). A skilled artisan is aware that sequences having N-terminal truncations preferably have an ATG start codon included.

#### A. Expression in a Prokaryote

[00204] The desired plasmids were prepared by ligating the mutated gene insert into the similarly digested vectors pET32c(+) (Novagen; Madison, WI) and pRS426GAL (Hua, 2000), a multiple copy yeast expression vector. These plasmids was expressed in *E. coli* BL21(DE3) (Novagen; Madison, WI). *E. coli* cells were grown in Luria-Bertani medium supplemented with 100 µg/mL ampicillin at 37 °C with shaking to OD<sub>600</sub> ~0.6. The following parameters were tested: isopropyl 1-thio-β-D-galactopyranoside (IPTG) concentration (50, 100, 250, 500, and 1000 µM); and temperature and time (20°C for 2, 3, 4, 6, 21 hours, 22 °C for 6, 8, 16, 19, 22, 45 hours, and 30°C for 3, 6 hours). The following assay conditions were tested to obtain maximum diterpene product yield: 30 mM HEPES (N-2-hydroxyethylpiperazine-N'-4-banesulfonic acid), pH 6.9, 7.2, 7.6, 8.0; 30 mM Tris (tris(hydroxymethyl)aminomethane hydrochloride), pH 7.4, 7.8, 8.2; 1, 5, 10% glycerol; 1, 3, 5, 10, 20 mM DTT (dithiothreitol); 20 mM β-mercaptoethanol; 2, 5, 8% Triton X-100; 5% Tween 80; 0, 2,

7.5, 20, 50 mM MgCl<sub>2</sub>; 0, 30, 500, 1000 μM MnCl<sub>2</sub>; 2, 10, 13.3, 20, 40, 80, 200 μM GGDP; and 23°C and 32°C assay temperatures.

**[00205]** Optimal soluble protein production and diterpene yield were obtained with the following conditions. Cell cultures were induced with 1 mM IPTG at 20°C with shaking for 6 hours and lysed by sonication in 30 mM HEPES, pH 7.2, 5 mM DTT, and 5% glycerol. The soluble fraction of the lysate (100 mg/mL) was incubated with 20 μM GGDP in 30 mM HEPES, pH 7.2, 5 mM DTT, 5% glycerol, 2 mM MgCl<sub>2</sub>, and 500 μM MnCl<sub>2</sub> overnight at 32°C.

**[00206]** Levopimaradiene synthase was also tested for activity towards 200 μM geranyl diphosphate and farnesyl diphosphate. Cell cultures were induced and lysed as noted above. The soluble fraction of the lysate was incubated overnight at 32°C with 200 μM substrate in 30 mM HEPES, pH 7.2, 5 mM DTT, 5% glycerol, 2 mM MgCl<sub>2</sub> and 500 μM MnCl<sub>2</sub>, and overlaid with hexane (1 mL).

## **B. Expression in a Eukaryote**

**[00207]** Expression in *S. cerevisiae* JBY575, which represents wild-type yeast, was observed. JBY575 cells transformed with pRS426GAL inserted with the putative levopimaradiene synthase were grown in synthetic complete medium lacking uracil and supplemented with 2% glucose and 1.5% agar at 30°C to saturation and induced with galactose for 48 hours. Cells were harvested, resuspended in lysis buffer, and mixed by vortexing over glass beads. The lysate was assayed with 60 μM GGDP in the presence and absence of 0.2% and 5% Triton X-100.

**[00208]** All *in vitro* reactions were extracted 3x with hexane and dried over MgSO<sub>4</sub>. The reaction was further extracted twice with diethyl ether and dried over MgSO<sub>4</sub>. Thereafter, the crude lysate was suspended in 100 mM Tris, pH 8.0 containing 2.9 units/mg apyrase (a dephosphorylating agent) and 10 units/μL calf intestinal alkaline phosphatase, incubated at 30°C for 3 hours, and extracted with diethyl ether as noted above (Croteau and Cane, 1985). The crude reaction mixtures were eluted over SiO<sub>2</sub>, concentrated, and analyzed by GC and GC/MS.

**[00209]** Gas chromatography spectra were obtained on a Hewlett Packard 6890 Series GC System equipped with an Rt<sub>x</sub>-5 capillary column (Restek, 30 m x 0.25 mm i.d., 0.10 μm d<sub>f</sub>). The following separation conditions and temperature program were

employed: injector port 250 °C, FID 250°C, split ratio 39:1, helium flow 20 cm/s (0.6 mL/min), 150°C hold 5 min, increase to 250°C (5 °C/min), hold 5 min. GC/MS spectra were obtained on a Hewlett-Packard 5890A instrument with a 30-m DB-5ms column (J&W Scientific Inc., 0.25 mm i.d., 0.10 µm df). The following separation conditions and temperature program were employed: injector port 280°C, transfer lines: 280°C, helium flow at 30 cm/s (1 mL/min) with splitless injection at 150°C hold 3 min, increase to 250°C (5 °C/min), hold 5 min. Mass spectra ( $m/z$  50 to 500) were obtained on a VG ZAB-HF reverse-geometry double-focusing instrument at 70 eV with an electron-impact ion source (200°C). Accelerating voltage was set to 8 kV and the resolution was 1000 (10% valley).

## EXAMPLE 5

### LEVOPIMARADIENE STANDARD-SYNTHESIS AND STRUCTURAL CONFIRMATION

**[00210]** Levopimarol (95.0 mg, 0.33 mmol) was dissolved in 3.7 mL dichloromethane and 92 µL triethylamine then cooled to 0°C. Methanesulfonyl chloride (31 µL, 0.39 mmol) was added dropwise *via* syringe. The reaction was monitored by thin layer chromatography (TLC) (1:1 chloroform:diethyl ether) and quenched after 15 min with ice-cold saturated aqueous sodium bicarbonate. The solution was extracted with dichloromethane (3 x), washed with H<sub>2</sub>O, dried with MgSO<sub>4</sub>, filtered, and concentrated. (Cambie *et al.*, 1990) The crude material was purified by preparative TLC (1 mm SiO<sub>2</sub>, 1:1 chloroform:diethyl ether) yielding 75.1 mg levo-8,12-dien-18-yl methanesulfonate (62.4% yield, R<sub>f</sub> 0.96). <sup>1</sup>H NMR (CDCl<sub>3</sub>, 400 MHz) δ 5.54 (q,  $J$  = 1.8 Hz, 1H, H-14), 5.15 (t,  $J$  = 4.3 Hz, 1H, H-12), 3.97 (d,  $J$  = 9.4 Hz, 1H, H-18), 3.73 (d,  $J$  = 9.4 Hz, 1H, H-18), 3.00 (s, 3H, CH<sub>3</sub>SO<sub>2</sub>), 2.38-2.28 (m, 3H, H-7α, H-11α, H-11β), 2.19-2.03 (m, 3H), 1.76 (dt, 1H), 1.61-1.52 (m, 3H), 1.47-1.34 (m, 4H), 1.27-1.21 (m, 1H), 0.97 (d, 6H, H-16, H-17), 0.91 (s, 3H), 0.88 (s, 3H).

**[00211]** Under an inert atmosphere, the mesylate (21.5 mg, 0.06 mmol) was dissolved in tetrahydrofuran in a Schlenk flask equipped with a cold finger. Excess lithium triethylborohydride (263 µL, 1 M in tetrahydrofuran, 0.26 mmol) was added dropwise to the solution, the reaction was stirred at reflux for 6 hours and monitored by TLC (6:1 hexane:diethyl ether). The reaction was quenched with ice-cold H<sub>2</sub>O, extracted with hexane (3x), dried with MgSO<sub>4</sub>, filtered over a silica plug, and concentrated. (Walter, 1988) GC analysis indicated that an 85:15 mixture of levopimaradiene:abietatriene had

been obtained in 41.3% yield (6.6 mg). The isomeric mixture was separated by argentic TLC (Li *et al.*, 1995) (SiO<sub>2</sub>-AgNO<sub>3</sub>, 3 developments with 85:15 hexane:diethyl ether) giving pure abietatriene (R<sub>f</sub> 0.96) (Kutney and Han, 1996) and pure levopimaradiene (R<sub>f</sub> 0.92) as identified by *de novo* characterization based on <sup>1</sup>H, COSY-DEC, <sup>13</sup>C, DEPT-135, HSQC, and HMBC NMR, and GC/MS analyses.

**[00212]** <sup>1</sup>H NMR (CDCl<sub>3</sub>, 500 MHz, 25 °C) δ 5.518 (q, *J* = 1.8 Hz, 1H, H-14), 5.141 (br tq, *J* = 4.3, 1.3 Hz, 1H, H-12), 2.338 (ddd, *J* = 13.3, 4.5, 2.2 Hz, 1H, H-7β), 2.323, 2.307 (m, 2H, H-11α, H-11β), 2.145 (septet of q, *J* = 6.8, 1.3 Hz, 1H, H-15), 2.075 (br td, *J* = 13.2, 5.2 Hz, 1H, H-7α), 2.021 (ddt, *J* = 11.5, 8.6, 1.8 Hz, 1H, H-9α), 1.737 (dtd, *J* = 12.9, 3.4, 1.6 Hz, 1H, H-1β), 1.697 (ddt, *J* = 12.7, 5.4, 2.7 Hz, 1H, H-6α), 1.519 (dt, *J* = 13.5, 3.4 Hz, 1H, H-2α), 1.444 (m, 1H, H-2β), 1.386 (m, 1H, H-3β), 1.368 (qd, *J* = 12.8, 4.5 Hz, 1H, H-6β), 1.149 (tdd, *J* = 13.2, ~4.0, 0.8 Hz, 1H, H-3α), 1.045 (dd, *J* = 12.5, 2.8 Hz, 1H, H-5α), 0.975 (d, *J* = 6.8 Hz, 6H, H-16, H-17), 0.866 (td, *J* = ~12.7, ~3.4 Hz, 1H, H-1α), 0.862 (s, 3H, H-18), 0.861 (s, 3H, H-20), 0.821 (s, 3H, H-19). Chemical shifts were referenced to Si(CH<sub>3</sub>)<sub>4</sub> and are accurate to ± 0.001. Coupling constants are accurate to ± 0.5 Hz.

**[00213]** <sup>13</sup>C NMR (CDCl<sub>3</sub>, 125 MHz, 25 °C) δ 139.46 (C-8), 138.91 (C-13), 118.73 (C-14), 114.87 (C-12), 55.23 (C-5), 49.61 (C-9), 42.21 (C-3), 40.75 (C-10), 37.91 (C-1), 36.15 (C-7), 33.48 (C-20), 33.45 (C-4), 33.26 (C-15), 23.80 (C-6), 22.75 (C-11), 21.80 (C-19), 21.45 (C-17), 21.37 (C-16), 19.00 (C-2), 14.10 (C-18). Chemical shifts (±0.2 ppm) were referenced to the CDCl<sub>3</sub> signal at 77.0 ppm.

**[00214]** GC/MS EI<sup>+</sup> *m/z* (%) = 272 [M<sup>+</sup>] (73), 257 [M - CH<sub>3</sub>] (13), 229 [M - CH(CH<sub>3</sub>)<sub>2</sub>] (7), 148 [M - C<sub>9</sub>H<sub>16</sub>] (64), 147 (27), 146 (50), 137 [M - C<sub>10</sub>H<sub>15</sub>] (94), 136 (65), 135 (60), 134 (90), 133 (66), 131 (43), 119 (27), 117 (34), 105 (58), 95 (31), 93 (28), 92 (100), 91 [M - C<sub>13</sub>H<sub>25</sub>] (97), 83 (20), 81 (26), 69 (24). GC co-elution of synthetic levopimaradiene with the enzymatic product and GC/MS fragmentation confirmed the identity of the *in vitro* product diterpene as levopimaradiene.

## EXAMPLE 6

### ISOLATION AND CHARACTERIZATION OF A DITERPENE CYCLASE cDNA

#### FROM *GINKGO BILOBA*

[00215] *G. biloba* cDNA libraries were prepared from cultivated embryonic roots. A homology-based approach utilizing PCR was employed to screen the library. Degenerate primers were designed based on conserved sequence regions among gymnosperm terpene synthases. These included *Abies grandis* abietadiene synthase, a bifunctional diterpene synthase that directs both proton-initiated cyclization and ionization of the divalent metal cation-diphosphate ester moiety; and synthases that effect diphosphate ionization to induce cyclization, including the diterpene *Taxus brevifolia* taxadiene synthase; the sesquiterpenes *Abies grandis* bisabolene synthase, selinene synthase, and humulene synthase; and the monoterpenes *Abies grandis* pinene synthase, limonene synthase, and myrcene synthase. Seven forward and eight reverse degenerate primers identifying eight regions of high sequence homology were designed. The combination of HSG1FP with HSG6RP resulted in amplification of a 1907 bp fragment (HSG97), which was determined to have significant sequence homology to higher plant terpene cyclases. A segment of this fragment was <sup>32</sup>P-labeled and used as a hybridization probe for high stringency screening of 3.0 x 10<sup>5</sup> colonies from cDNA library HSG2. A skilled artisan is aware that the cDNA preferably comprises a majority of expressed sequences, which are also preferably full-length, from an organism.

[00216] A total of 10 hybridizing colonies were obtained and put through a secondary round of high stringency screening producing an enriched pool of clones. The termini of the three longest cDNAs were sequenced and identified as putative diterpene cyclases based on homology to *Abies grandis* abietadiene synthase. Furthermore, all three clones had identical 5' and 3' ends (approximately 600 bp at each end). One clone was further characterized. Sequence data revealed that a 2681 bp cDNA had been cloned, however, the initiation codon was absent. RNA ligase mediated rapid amplification of cDNA ends ("RACE") was employed to isolate the 5'-untranslated region and the methionine start site. The full-length gene, *Ginkgo biloba* levopimaradiene synthase, a diterpene synthase, was 2705 bp in length and encoded an 873 amino acid open reading frame of 2622 bp with a predicted molecular weight of 100,289 (see FIG. 4).

## EXAMPLE 7

### ANALYSIS AND SELECTION OF PLASTID TARGETING SEQUENCE

#### CLEAVAGE SITE

[00217] Cytosolically synthesized plastid proteins contain N-terminal targeting sequences that direct their translocation to specific plastidial compartments. Proteolysis of the signal sequence occurs by a specific protease, yielding the mature protein. Plastid transit peptides typically range between 30 to 80 amino acids in length; are rich in hydroxylated amino acids, basic amino acids, and small hydrophobic residues; and display low contents of tyrosine and acidic residues. For purposes of heterologous expression, wherein native processing peptidases are not present, cleavage of the signal sequence may be required prior to expression to avoid formation of inclusion bodies. In general, cleavage sites are distinguished by a decreased frequency of serine residues and a corresponding increase in the frequency of tyrosine and acidic amino acids. In a majority of higher eukaryotes, arginine is found at positions -2 and -6 to -10 relative to the cleavage site. Furthermore, a consensus motif of (Val/Ile)-X-(Ala/Cys)↓Ala (wherein the downward arrow (↓) indicates the site of bond cleavage) has been identified in a series of stroma-targeting chloroplast transit peptides with known cleavage sites (von Heijne and Nishikawa, 1991; von Heijne and Gavel, 1990; Keegstra and Olsen, 1989).

[00218] Analysis of *Ginkgo biloba* levopimaradiene synthase indicated the following representation of amino acid residues: the first tyrosine residue at Y<sup>84</sup>; the first glutamic acid residue at E<sup>64</sup>; the first aspartic acid residue at D68; and a decreased frequency of serine residues between S<sup>47</sup> and S<sup>96</sup>. Two potential cleavage sites were identified at Ile-His-Ala<sup>60</sup>↓Ser<sup>61</sup> (with arginine at -9 and -11 relative to the cleavage site) and at Ile-Gln-Cys<sup>127</sup>↓Met<sup>128</sup> (with arginine at -11 relative to the cleavage site). Submission of *Ginkgo biloba* levopimaradiene synthase to META Predict Protein Chloro P predicted the presence of an N-terminal chloroplast transit peptide with a cleavage site between H<sup>59</sup>-A<sup>60</sup> (Nielsen *et al.*, 1995). Three truncation sites were selected in consideration of the data presented above: Ala<sup>60</sup>-Ser<sup>61</sup> (hereafter referred to as Ser<sup>61</sup>),



Arg<sup>79</sup>-Leu<sup>80</sup> (hereafter referred to as Leu<sup>80</sup>), and Cys<sup>127</sup>-Met<sup>128</sup> (hereafter referred to as Phe<sup>129</sup>).

[00219] Recently, successful heterologous expression of truncated levopimaradiene synthases have been reported. Cleavage of the N-terminal 84 residues of *Abies grandis* abietadiene synthase produced active protein (Ravn *et al.*, 2000). Truncation of 79 or fewer residues of *Taxus brevifolia* taxadiene synthase produced functional protein, however, elimination of 93 or more residues resulted in loss of catalytic activity (Williams *et al.*, 2000). Low primary sequence homology is observed between *Ginkgo biloba* levopimaradiene synthase, *Abies grandis* abietadiene synthase, and *Taxus brevifolia* taxadiene synthase prior to residue *Ginkgo biloba* levopimaradiene synthase Trp<sup>89</sup>, 21.5% and 14.0%, respectively. However, significant sequence similarity begins at position Trp<sup>89</sup>, 65.7% and 44.1% respectively. Furthermore, no distinct identity is apparent between these synthases at the truncation sites reported to produce functional protein.

## EXAMPLE 8

### LEVOPIMARADIENE SYNTHASE SEQUENCE ANALYSIS

[00220] Protein analysis of the deduced polypeptide indicated *Ginkgo biloba* levopimaradiene synthase to have high sequence similarity to *Abies grandis* abietadiene synthase (60%), *Abies grandis* bisabolene synthase (46%), and *Taxus brevifolia* taxadiene synthase (41%). Three aspartate-rich motifs and a putative plastidial transit peptide were identified in *Ginkgo biloba* levopimaradiene synthase. An N-terminal DDXID motif (*Ginkgo biloba* levopimaradiene synthase 91-95), also observed in *Abies grandis* abietadiene synthase, may serve to stabilize carbocations and/or direct deprotonation. Crystallographic and mutagenesis studies suggest that the consensus motif, D(I/V)DDTA (*Ginkgo biloba* levopimaradiene synthase 405-410), initiates cyclization of GGDP. Moreover, this aspartate-rich sequence remains highly conserved among synthases that effect proton-initiated cyclization, including copalyl diphosphate synthases, *Abies grandis* abietadiene synthase, *Phaeosphaeria ent-kaurene* synthase, and squalene-hopene cyclases (Bohlman *et al.*, 1998). A carboxy-terminal DDXXD motif (*Ginkgo biloba* levopimaradiene synthase 624-628) resides in kaurene synthases; *Abies grandis* abietadiene synthase; prenyltransferases; and in plant mono-, sesqui-, and diterpene

synthases (Bohlman *et al.*, 1998). This domain in a specific embodiment affects binding of the divalent metal ion-diphosphate complex. Crystal structure analysis of tobacco epiaristolochene synthase identified two  $Mg^{2+}$  ions bound at the entrance of the active site by coordination to aspartic acid residues of the DDXXD motif (Starks *et al.*, 1997).

[00221] Comparative protein analysis indicated that *Ginkgo biloba* levopimaradiene synthase contained features reminiscent of two distinct catalytic domains, and thereby confirmed it as a bifunctional levopimaradiene synthase. Furthermore, *Ginkgo biloba* levopimaradiene synthase displayed a high degree of homology to conserved amino acid residues of mono-, sesqui-, and diterpene secondary metabolite families (Bohlman *et al.*, 1997). These included the absolutely or highly conserved residues Ser<sup>459</sup>, Ala<sup>472</sup>, Pro<sup>713</sup>, Cys<sup>789</sup>, Arg<sup>414, 417, 587, 610, 766</sup>, the acidic residues Asp<sup>363, 624, 625, 628, 770, 851</sup> and Glu<sup>466, 567, 592, 703, 717</sup>, and the aromatic residues His<sup>419</sup>, Phe<sup>431, 438, 585, 663</sup>, Tyr<sup>523, 594, 700, 847</sup>, and Trp<sup>89, 574, 646, 706</sup>. Three significant deviations in the *Ginkgo biloba* levopimaradiene synthase sequence included a highly conserved histidine which corresponds to *Ginkgo biloba* levopimaradiene synthase Tyr<sup>373</sup>, an absolutely conserved proline which corresponds to *Ginkgo biloba* levopimaradiene synthase Arg<sup>655</sup>, and an absolutely conserved acidic amino acid which corresponds to *Ginkgo biloba* levopimaradiene synthase Gly<sup>672</sup>.

## EXAMPLE 9

### PROTEIN EXPRESSION AND OPTIMIZATION OF *IN VITRO* ENZYMATIC ACTIVITY

[00222] Site-directed mutagenesis was employed to insert a *Sall* site followed by seven adenines and a methionine start codon at Ala<sup>2</sup>, Ser<sup>61</sup>, Leu<sup>80</sup>, and Phe<sup>129</sup>. Following removal of the 5'-untranslated region and plastid targeting sequence, the desired plasmids were prepared by ligation with pET32c(+) (a bacterial expression system containing a thioredoxin tag designed for maximal production of soluble protein) and pRS426GAL (a multiple copy yeast expression system). The levopimaradiene synthase was expressed in the *E. coli* strain BL21(DE3) and the wild-type *S. cerevisiae* strain JBY575, respectively. *E. coli* cells were grown in Luria-Bertani medium supplemented with ampicillin and induced with IPTG. SDS-PAGE analysis indicated that protein production increased with time and reached maximum accumulation by 21 hours. However, recombinant protein resided mainly in the insoluble fractions of the lysate,

indicating that it was likely encapsulated in an inclusion body. Attempts to improve protein solubility by variation of IPTG concentrations between 50 to 1000 mM were unsuccessful. However, employing an induction temperature of 20°C produced functionally soluble protein.

**[00223]** The skilled artisan recognizes that lysis and assay conditions should be optimized. Examples of parameters that can be adjusted to optimize conditions include altering buffer, pH, reductant, reductant concentration, metal cofactors, cofactor concentrations, glycerol concentrations, substrate concentrations, and assay temperatures. Levopimaradiene synthase activity proved to be independent of magnesium but required manganese cofactor for catalysis. Maximum soluble protein production and diterpene yield were obtained with the following conditions. Cell cultures were induced with 1 mM IPTG at 20°C with shaking for 6 hours and lysed by sonication in 30 mM HEPES, pH 7.2, 5 mM DTT, and 5% glycerol. The soluble fraction of the lysate (100 mg/1 mL) was incubated with 20  $\mu$ M GGDP in 30 mM HEPES, pH 7.2, 5 mM DTT, 5% glycerol, 2 mM  $MgCl_2$ , and 500  $\mu$ M  $MnCl_2$  overnight at 32°C. Yeast expression was induced with galactose at 30°C for 48 hours and the cells were lysed and assayed according to the conditions noted.

**[00224]** The extent of N-terminal truncation affected catalytic activity in both expression hosts. Bacterial expression of *Ginkgo biloba* levopimaradiene synthase truncated at Ala<sup>2</sup>, Ser<sup>61</sup>, and Leu<sup>80</sup> produced levopimaradiene as the exclusive diterpene hydrocarbon, however, Phe<sup>129</sup> failed to produce detectable levels of any diterpene. Highest expression levels were obtained with the Ser<sup>61</sup> truncation (approximately 1% turnover of GGDP) and lowest levels were obtained with Ala<sup>2</sup>, with approximately 80% difference in activity. Yeast expression of Ser<sup>61</sup> and Leu<sup>80</sup> yielded levopimaradiene as the sole diterpene. However, both the Ala<sup>2</sup> and Phe<sup>129</sup> truncated genes failed to produce observable levels of any diterpene product. Controls performed in parallel did not yield levopimaradiene. *E. coli* expression and incubation with geranyl diphosphate and farnesyl diphosphate did not produce any identifiable terpenes by GC or GC/MS. However, a skilled artisan is aware of parameters which may be optimized and/or additional sequences which may be employed to detect and/or increase synthesis of levopimaradiene synthase.

## EXAMPLE 10

### PRODUCT CHARACTERIZATION

[00225] Due to the low levels of levopimaradiene production, a synthetic standard to confirm product identification was utilized. Levopimaric acid was obtained from Helix Biotech and converted to levopimarol according to literature procedures, with care employed to minimize exposure to oxygen and heat. Reaction of the alcohol with mesyl chloride and triethylamine followed by silica gel purification resulted in a 62% yield of the mesylate derivative. The ester was reduced with Super-Hydride® to yield a 41% mixture of levopimaradiene:abietatriene (85:15). Argentic chromatography effected the separation of the hydrocarbons; levopimaradiene was identified by NMR and GC/MS analysis. Co-elution on GC and identical GC/MS fragmentation of the biosynthetic hydrocarbon with the synthetically prepared levopimaradiene confirmed identification of the enzyme product to be levopimaradiene.

[00226] Molecular biology techniques were employed to confirm the presence of levopimaradiene as a rapidly metabolized intermediate. The bifunctional enzyme directs a multi-step mechanistic sequence in which GGDP is cyclized to labdadienyl diphosphate which undergoes allylic ionization of the ester moiety followed by hydrogen shift, methyl migration, and deprotonation yielding levopimaradiene.

## EXAMPLE 11

### EXPRESSION *IN VIVO* IN YEAST THAT ACCUMULATE GGDP

[00227] A skilled artisan recognizes that it is preferable to increase the amount of substrate provided for in the production of levopimaradiene to ultimately increase ginkgolide yields. Thus, the invention preferably includes an increase in the amount of effective geranylgeranyl diphosphate, which is upstream in the ginkgolide biosynthetic pathway. A skilled artisan recognizes that an effective amount of geranylgeranyl diphosphate is that which is subject to metabolism in the isoprene biosynthetic pathway. A skilled artisan is aware that an increase in GGDP occurs in multiple ways, such as by providing GGDP exogenously or by increasing its production through transgenic and/or bioengineering means. GGDP is increased by the methods and compositions of the invention described in the U.S. Patent Application entitled,

"Diterpene-Producing Unicellular Organism" filed on the same day and incorporated by reference herein.

**[00228]** The yeast strain of the copending application employed was EHY18 (Hart, E., 2001) which was further transformed with multiple-copy yeast expression vectors comprising an isolated and purified nucleic acid sequence of GbLS or derivatives thereof, which are described herein under control of the GAL1 inducible promoter. Yeast cells were grown to saturation in 5 mL ScD-Leu-Ura at 30 °C. Cells were harvested (1300 × g, 2 min, 25 °C) and resuspended in 500 µL sterile Milli-Q H<sub>2</sub>O (2 ×). The washed cells were resuspended in 5 mL sterile Milli-Q H<sub>2</sub>O. An aliquot of the culture (5 µL) was added to a 25 mL Corex tube containing pre-prepared 2 × resin and 5 mL ScG-Leu-Ura (4% galactose) and shaken at 30°C for 6 days. Resin was collected by filtration through a Kontes chromatography column (1.7 cm diameter) and rinsed with dI H<sub>2</sub>O until eluent was clear. The resin was incubated in 2 mL ethanol for several minutes, eluted, and repeated 2 ×. The combined eluents were dissolved in approximately 3 mL dI H<sub>2</sub>O and extracted (3 ×) with 3 mL hexane. The combined extracts were concentrated under a stream of nitrogen. GC analysis (quantitation extrapolated from 0.2 mg/mL longifolene external standard) indicated the following product profile (obtained from triplicate measurement of three cultures): 0.61 ± 0.20 mg/L levopimaradiene, 0.04 ± 0.01 mg/L abietadiene, 0.15 ± 0.05 mg/L abietatriene, 0.16 ± 0.13 mg/L (+)-copalol, and 1.79 ± 0.90 mg/L geranylgeraniol. GC/MS fragmentation confirmed the identity of each compound.

**[00229]** A large scale culture (1-L) was prepared by growing the yeast cells to saturation in 5 mL ScD-Leu-Ura at 30 °C. Cells were harvested (1300 × g, 2 min, 25 °C) and resuspended in 500 µL sterile Milli-Q H<sub>2</sub>O (2 ×). The washed cells were resuspended in 5 mL sterile Milli-Q H<sub>2</sub>O. An aliquot of the culture (1 mL) was added to a 2 L flask containing 1-L ScG-Leu-Ura (4% galactose) and 88.27 g resin in 120 mL Milli-Q H<sub>2</sub>O (pre-prepared). The resin was eluted with ethanol (3 × 100 mL with 20 min incubation periods), which was subsequently extracted (6 ×) with hexane. The combined extracts were dried over magnesium sulfate, filtered, and concentrated under reduced pressure. GC analysis (quantitation extrapolated from 0.2 mg/mL longifolene external standard) indicated the following product profile (obtained from triplicate measurement): 0.29 ± 0.04 mg/L levopimaradiene, 0.04 ± 0.02 mg/L abietadiene, 0.13 ± 0.04 mg/L

abietatriene,  $0.11 \pm 0.02$  mg/L (+)-copalol, and  $0.56 \pm 0.03$  mg/L geranylgeraniol. GC/MS analyses confirmed the production of the above noted diterpenes, which included hydrocarbons and an alcohol.

## EXAMPLE 12

### OPTIMIZATION OF LEVOPIMARADIENE PRODUCTION

[00230] Similar to levopimaradiene synthase activity in *E. coli*, catalytic activity in EHY18 was affected by the N-terminal truncation of GbLS. Expression of wild-type levopimaradiene synthase and its truncated counterparts Ser61 and Leu80 produced (+)-copalol, levopimaradiene, abietadiene, and abietatriene. Co-elution on GC and identical GC/MS fragmentation of the biosynthetic alcohol with synthetic (+)-copalol allowed the unequivocal identification of the enzyme product as (+)-copalol. GC analysis of the expressed GbLS Phe129 indicated no diterpene alcohol or hydrocarbon formation. Highest diterpene production was obtained with the Ser61 and Leu80 truncated constructs (Ser61  $\geq$  Leu80) and was approximately four times greater than that observed for the wild-type synthase. Negative controls performed in parallel (expressing EHY18) did not yield levopimaradiene, abietadiene, abietatriene, or (+)-copalol.

[00231] Further optimization studies employed the Ser61 construct. Factors influencing diterpene production included induction period and concentration of galactose and resin in the inducing medium. Highest diterpene yields were observed with 4% galactose and 0.70 g resin/5 mL culture medium. Expression for 6 days, in the above noted induction medium, resulted in a net yield of  $\sim 0.8$  mg/L diterpene hydrocarbons and  $\sim 0.2$  mg/L (+)-copalol, as determined by GC quantitation.

[00232] Expression of GbLS in EHY18 resulted in a three- to six-fold increase in levopimaradiene yield relative to the bacterial expression system previously employed. In addition, expression of GbLS in EHY18 afforded abietatriene, the immediate hydrocarbon precursor of the ginkgolides. The ability to enhance levopimaradiene production in a yeast cell having an increased effective amount of geranylgeranyl diphosphate suggests the system is useful in the isolation of the first oxygenase involved in ginkgolide biosynthesis. Lastly, identification of (+)-copalol represents the first observation of the free intermediate of (+)-copalyl pyrophosphate by a bifunctional diterpene catalyst and supports previous data implicating (+)-copalyl

pyrophosphate as a precursor to levopimaradiene (Peters *et al.*, 2000; Schwarz and Arigoni, 1999).

**[00233]** Bacterial expression optimization includes standard manipulations known in the art to overcome problems such as low solubility and low expression levels. For instance, a skilled artisan is aware that different *E. coli* expression systems, including commercially available vectors and strains, are utilized to produce higher amounts of soluble protein. An example of a vector includes pSBET (Schenk *et al.*, 1995), which is particularly useful for heterologous expression in *Escherichia coli* of plant genes that often have a significant number of arginine residues. The vector is particularly well-suited to *Escherichia coli* BL21 (DE3) (Sambrook *et al.*, 1989). Also, many different *E. coli* strains are known in the art and may be used, such as LE392 cells, DH5 $\alpha$  cells, or SURE<sup>TM</sup> (Stratagene; La Jolla, CA) cells.

### EXAMPLE 13

#### CLONING GINKGOLIDE BIOSYNTHETIC GENES

**[00234]** Difficulties in RNA extraction from recalcitrant gymnosperm tissue have been noted (Chang, *et al.*, 1993; Lewinsohn, *et al.*, 1994). High levels of polysaccharides in gymnosperm tissue and oxidation of polyphenols during extraction resulted in contaminated and/or degraded RNA. With respect to *G. biloba* tissue, studies indicate that mature ginkgo seeds are comprised of approximately 35% water-soluble polysaccharides (Arahira, *et al.*, 1994). However, successful construction of a *Ginkgo biloba* cDNA library, as described herein, has overcome this problem.

**[00235]** In a specific embodiment, nucleic acid sequences encoding other enzymes in the ginkgolide biosynthesis pathway are obtained. In a specific embodiment, a cDNA library, such as for *E. coli* or *S. cerevisiae*, comprising *Ginkgo biloba* sequences are exposed to an *E. coli* or *S. cerevisiae* cell, respectively, wherein the cell also comprises the levopimaradiene synthase sequence, and the presence of a desired downstream product is assayed. In a specific embodiment, the GC and/or GC/MS profile of the product is known and its presence is determined. In a further specific example, the nucleic acid sequence for a dehydrogenase, which generates formation of abietatriene, is cloned by assaying pools of cells harboring levopimaradiene synthase and identifying by chromatography (*i.e.*, GC or GC/MS) the pool in which abietatriene is produced. Once a pool is identified, this pool is

broken down into its constituents which are assayed in smaller pools and/or individually to identify the cell containing the clone expressing the desired nucleic acid sequence.

[00236] In an embodiment of the present invention, a first ginkgolide biosynthetic gene downstream of levopimaradiene synthase is provided in a cell comprising the levopimaradiene synthase, wherein both the first downstream gene and the levopimaradiene synthase are expressed concomitantly. In a specific embodiment, the cell provides biosynthesis of a ginkgolide biosynthetic intermediate that is a first derivative of levopimaradiene, such as abietatriene. A further embodiment is the subsequent cloning of a second downstream ginkgolide biosynthetic gene, which allows biosynthesis of a different ginkgolide biosynthetic intermediate upon expression in a cell comprising the levopimaradiene synthase, the first downstream ginkgolide biosynthetic gene and the second ginkgolide biosynthetic gene; this cell demonstrates biosynthesis of a ginkgolide biosynthetic intermediate that is a second derivative of the first derivative (e.g., abietatriene) of levopimaradiene. Other such embodiments are contemplated in which levopimaradiene, produced by a cell that expresses the *Ginkgo biloba* levopimaradiene synthase and conservatively modified variants, serves as an intermediate in biosynthesis of a diterpenoid, and preferably a ginkgolide.

#### EXAMPLE 14

##### SUMMARY OF THE PRESENT INVENTION

[00237] Levopimaradiene synthase, which directs the first committed step in ginkgolide biosynthesis, was cloned and functionally characterized as part of a program to isolate and express genes involved in the biosynthesis of the ginkgolides. A *Ginkgo biloba* cDNA library was prepared from embryonic roots and screened utilizing a homology-based approach employing degenerate primers with high sequence similarity to gymnosperm terpene synthases. Polymerase chain reaction amplification provided a 1907 bp fragment, which was employed to probe the library. Colony hybridization and rapid amplification of cDNA ends yielded a full-length clone with a 2622 bp open reading frame encoding a predicted protein sequence of 873 amino acids with an estimated molecular weight of 100,289. Protein analysis indicated that a bifunctional terpene cyclase had been isolated with high sequence identity to *Abies grandis* abietadiene synthase (60%), *Abies grandis* bisabolene synthase (46%), and *Taxus brevifolia* taxadiene synthase (41%).



Additionally, the amino acid sequence contained a putative N-terminal plastidial transit peptide and three aspartate-rich regions.

**[00238]** Functional expression in *Escherichia coli* of the full-length cDNA and corresponding truncations at Ser<sup>61</sup> and Leu<sup>80</sup> provided enzymatic activity capable of cyclizing geranylgeranyl diphosphate to levopimaradiene, as confirmed by GC/MS analysis. Expression of the truncated Phe<sup>129</sup> gene product resulted in complete loss of enzymatic activity. Functional expression in wild-type *Saccharomyces cerevisiae* of the Ser<sup>61</sup> and Leu<sup>80</sup> truncations yielded levopimaradiene synthase activity, albeit in lower yields than with the bacterial system, whereas the full-length and Phe<sup>129</sup> clones failed to produce detectable levels of biosynthetic product. Isolation and characterization of levopimaradiene synthase represents the first confirmation of an enzyme involved in ginkgolide biosynthesis.

**[00239]** An engineered yeast strain has been employed to achieve increased levopimaradiene production levels. An approximate three-fold to six-fold increase in levopimaradiene yield was obtained relative to the previously employed bacterial and yeast expression systems. In addition, production of abietatriene, the direct hydrocarbon progenitor of the ginkgolides, was realized.

## REFERENCES

**[00240]** All patents and publications mentioned in the specification are indicative of the levels of those skilled in the art to which the invention pertains. All patents and publications are herein incorporated by reference to the same extent as if each individual publication was specifically and individually indicated to be incorporated by reference herein.

## PATENTS

U.S. Patent No. 5,302,523, issued April 12, 1994.

U.S. Patent No. 5,322,783, issued June 21, 1994.

U.S. Patent No. 5,384,253, issued January 24, 1995.

U.S. Patent No. 5,464,765, issued November 7, 1995.

U.S. Patent No. 5,508,184, issued April 16, 1996.

U.S. Patent No. 5,538,877, issued July 23, 1996.

10041007-010702

U.S. Patent No. 5,538,880, issued July 23, 1996.  
U.S. Patent No. 5,550,318, issued August 27, 1996.  
U.S. Patent No. 5,563,055, issued October 8, 1996.  
U.S. Patent No. 5,591,616, issued January 7, 1997.  
U.S. Patent No. 5,610,042, issued March 11, 1997.  
U.S. Patent No. 5,840,873, issued Nov. 24, 1998  
U.S. Patent No. 5,843, 640, issued Dec. 1, 1998  
U.S. Patent No. 5,843,650, issued Dec. 1. 1998  
U.S. Patent No. 5,843,651, issued Dec. 1, 1998  
U.S. Patent No. 5,843,663, issued Dec. 1, 1998  
U.S. Patent No. 5,846,708, issued Dec. 8, 1998  
U.S. Patent No. 5,846,709, issued Dec. 8, 1998  
U.S. Patent No. 5,846,717, issued Dec. 8, 1998  
U.S. Patent No.5,846,726, issued Dec. 8, 1998  
U.S. Patent No. 5,846,729, issued Dec. 8, 1998  
U.S. Patent No. 5,846,783, issued Dec. 8, 1998  
U.S. Patent No. 5,849,481, issued Dec. 15, 1998  
U.S. Patent No. 5,849,483, issued Dec. 15, 1998  
U.S. Patent No. 5,849,486, issued Dec. 15, 1998  
U.S. Patent No. 5,849,487, issued Dec. 15, 1998  
U.S. Patent No. 5,849,497, issued Dec. 15, 1998  
U.S. Patent No. 5,849,546, issued Dec. 15, 1998  
U.S. Patent No. 5,849,547, issued Dec. 15, 1998  
U.S. Patent No. 5,851,770, issued Dec. 22, 1998  
U.S. Patent No. 5,851,772, issued Dec. 22, 1988  
U.S. Patent No.5,853,990, issued Dec. 29, 1998  
U.S. Patent No. 5,853, 993, issued Dec. 29, 1998  
U.S. Patent No. 5,853,992, issued Dec. 29, 1998  
U.S. Patent No. 5,856,092, issued Jan. 5, 1999  
U.S. Patent No. 5,858,652, issued Jan. 12, 1999  
U.S. Patent No. 5,861,244, issued Jan. 19, 1999  
U.S. Patent No.5,863,732, issued Jan. 26, 1999

10041007.010702

U.S. Patent No. 5,863,753, issued Jan. 26, 1999  
U.S. Patent No. 5,866,331, issued Feb. 2, 1999  
U.S. Patent No. 5,866,336, issued Feb. 2, 1999  
U.S. Patent No. 5,866,337, issued Feb. 2, 1999  
U.S. Patent No. 5,900,481, issued May 4, 1999  
U.S. Patent No. 5,905,024, issued May 18, 1999  
U.S. Patent No. 5,910,407, issued June 8, 1999  
U.S. Patent No. 5,912,124, issued June 15, 1999  
U.S. Patent No. 5,912,145, issued June 15, 1999  
U.S. Patent No. 5,912,148, issued June 15, 1999  
U.S. Patent No. 5,916,776, issued June 29, 1999  
U.S. Patent No. 5,916,779, issued June 29, 1999  
U.S. Patent No. 5,919,626, issued July 6, 1999  
U.S. Patent No. 5,919,630, issued July 6, 1999  
U.S. Patent No. 5,922, 574, issued July 13, 1999  
U.S. Patent No. 5,925,517, issued July 20, 1999  
U.S. Patent No. 5,925,525, issued Jul. 20, 1999  
U.S. Patent No. 5,928,862, issued July 27, 1999  
U.S. Patent No. 5,928,869, issued July 27, 1999  
U.S. Patent No. 5,928,870, issued, Jul. 27, 1999  
U.S. Patent No. 5,928,905, issued July 27, 1999  
U.S. Patent No. 5,928,906, issued July 27, 1999  
U.S. Patent No. 5,929,227, issued July 27, 1999  
U.S. Patent No. 5,932,413, issued Aug. 3, 1999  
U.S. Patent No. 5,932,451, issued Aug. 3, 1999  
U.S. Patent No. 5,935,791, issued Aug. 10, 1999  
U.S. Patent No. 5,935,825, issued Aug. 10, 1999  
U.S. Patent No. 5,939,291, issued Aug. 17, 1999  
U.S. Patent No. 5,942,391, issued Aug. 24, 1999  
European Application No. 320 308  
European Application No. 329 822  
GB Application No. 2 202 328

PCT Application No. PCT/US87/00880

PCT Application No. PCT/US89/01025

PCT Application WO 88/10315

PCT Application WO 89/06700

PCT Application WO 90/07641

PCT Application WO 92/17598

PCT Application WO 94/09699

PCT Application WO 95/06128

PCT Application WO 95/06128

#### NON-PATENTED LITERATURE

- [00241] Abad, A., Arno, M., Domingo, L. R., and Zaragoza, R. J. (1985) Tetrahedron 41, 4937-4940.
- [00242] Abdullah *et al.*, Biotechnology, 4:1087, 1986.
- [00243] Ayer, W. A., and Talamas, F. X. (1988) Can. J. Chem. 66, 1675-1685.
- [00244] Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Seidam, J.G., Smith, J. A., and Struhl, K. (1999) Curr. Prot. Mol. Biol. Wiley-Interscience, New York, NY.
- [00245] Balz, J.P., Courtois, D., Drieu, J., Drieu, K., Reynoard, J.P., Sohler, C., Teng, B.P., Touche, A., Petiard, V. (1999) Production of ginkgolides and bilobalide by *Ginkgo biloba* plants and tissue cultures. Planta Medica 65(7): 620-626.
- [00246] Bates, "Genetic transformation of plants by protoplast electroporation," Mol Biotechnol., 2(2):135-145, 1994.
- [00247] Batraw and Hall, "Stable transformation of sorghum-bicolor protoplasts with chimeric neomycin phosphotransferase II and beta glucuronidase genes," Theor. App. Genet., 82(2):161-168, 1991.
- [00248] Becker *et al.*, "Fertile transgenic wheat from microprojectile bombardment of scutellar tissue," Plant J., 5(2):299-307, 1994.
- [00249] Bhattacharjee; An; Gupta, J. Plant Bioch. and Biotech. 6, (2):69-73. 1997.

- 10041007, 010702
- [00250] Bohlman, J., Meyer-Gauen, G., and Croteau, R. (1998) Proc. Natl. Acad. Sci. USA 95, 4126-4133.
- [00251] Bower *et al.*, The Plant Journal, 2:409-416. 1992.
- [00252] Buising and Benbow, "Molecular analysis of transgenic plants generated by microprojectile bombardment: effect of petunia transformation booster sequence," Mol. Gen. Genet., 243:71-81, 1994.
- [00253] Callis *et al.*, Genes Dev., 1:1183-1200, 1987.
- [00254] Cambie, R. C., Rutledge, P. S., Ryan, G. R., Strange, G. A., and Woodgate, P. D. (1990) Aust. J. Chem. 43, 867-881.
- [00255] Cartayrade, A., Neau, E., Sohier, C., Balz, J-P.; Carde, J-P., and Walter, J. (1997) Plant Physiol. Biochem. 35, 859-868.
- [00256] Casas *et al.*, "Transgenic sorghum plants via microprojectile bombardment," Proc. Natl. Acad. Sci. USA, 90(23):11212-11216, 1993.
- [00257] Chang, S., Puryear, J., and Cairney, J. (1993) Plant Mol. Bio. Rep. 11, 113-116.
- [00258] Christou; Murphy; Swain, Proc. Nat'l Acad. Sci. U S A, 84(12):3962-3966, 1987.
- [00259] Coates, R.M., Ley, D. A., and Cavender, P. L. (1978) J. Org. Chem. 43, 4915-4922.
- [00260] Corey, E. J., Kim, C. U., and Takeda, M. (1972) Tetrahedron Lett. 42, 4339-4342.
- [00261] Corey, E. J., and Su, W-g. (1987) J. Am. Chem. Soc. 109, 7534-7536.
- [00262] Corey, E. J., Kang, M-c., Desai, M. C., Ghosh, A. K., and Houpiis, I. N. (1988) J. Am. Chem. Soc. 110, 649-651.
- [00263] Corey, E. J., and Ghosh, A. K.. (1988) Tetrahedron Lett. 29, 3205-3206.
- [00264] Corey, E. J., Matsuda, S. P. T., Baker, C. H., Ting, A. Y., Cheng, H. (1996) Biochem. Biophys. Res. Commun. 219, 327-331.

- [00265] Crimmins, M. T., Pace, J. M., Nantermet, P. G., Kim-Meade, A. S., Thomas, J. B., Watterson, S. H., and Wagman, A. S. (1999) *J. Am. Chem. Soc.* 121, 10249-10250.
- [00266] Croteau, R., and Cane, D. E. (1985) *Methods Enzymol.* 110, 383-405.
- [00267] Davisson, V. J., Woodside, A. B., and Poulter, C. D. (1985) *M. Enzymol.* 110, 130-144.
- [00268] Davisson, V. J., Woodside, A. B., Neal, T. R., Stremmler, K. E., Muehlbacher, M., and Poulter, C.D. (1986) *J. Org. Chem.* 51, 4768-4779.
- [00269] D'Halluin *et al.*, "Transgenic maize plants by tissue electroporation," *Plant Cell*, 4(12):1495-1505, 1992.
- [00270] Fiehe, K., Arenz, A., Drewke, C., Hemscheidt, T., Williamson, R.T., Leistner, E. (2000) Biosynthesis of 4'-O-Methylpyridoxine (Ginkgotoxin) from primary precursors. *J. Nat. Prod.* 63:185-189.
- [00271] Fromm *et al.*, *Nature*, 312:791-793, 1986.
- [00272] Fromm *et al.*, "Stable transformation of maize after gene transfer by electroporation," *Nature*, 319:791-793, 1986
- [00273] Gerhardt G., Rogalla K., Jaeger J. (1990) Drug therapy of disorders of cerebral performance. Randomized comparative study of dihydroergotoxine and Ginkgo biloba extract. *Fortschr Med* Jun 30;108(19):384-8.
- [00274] Ghosh-Biswas *et al.*, "Transgenic Indica rice (*Oryza sativa* L.) plants obtained by direct gene transfer to protoplasts," *J. Biotechnol.*, 32(1):1-10, 1994.
- [00275] Gigante, B., Santos, C, Fonseca, T., Curto, M. J. M., Luftmann, H., Bergander, K., and Berberan-Santos, M. N. (1999) *Tetrahedron* 55, 6175-6182.
- [00276] Hagio *et al.*, "Stable transformation of sorghum cell cultures after bombardment with DNA coated microprojectiles," *Plant Cell Rep.*, 10(5):260-264, 1991.
- [00277] Hart, E. In *Chemistry*, Rice University: Houston, 2001, p. 142.
- [00278] He *et al.*, *Plant Cell Reports*, 14 (2-3):192-196, 1994.
- [00279] Hensgens *et al.*, "Transient and stable expression of *gusA* fusions with rice genes in rice, barley and perennial ryegrass," *Plant Mol. Biol.*, 22(6):1101-1127, 1993.

- [00280] Hiei *et al.*, "Transformation of rice mediated by agrobacterium tumefaciens," Plant Mol. Biol., 35(1-2):205-218, 1997.
- [00281] Hill, A.M., Cane, D.E., Mau, C.J.D., West, C.A. High level expression of Ricinus communis castor oil synthase in *Escherichia coli* and characterization of the recombinant enzyme. (1996) Arch. Biochem. Biophys. 336(2): 283-289.
- [00282] Hosford, D. J., Domingo, M. T., Chabrier, P. E., and Braquet, P. (1990) Methods Enzymol. 187, 433-446.
- [00283] Hou and Lin, Plant Physiology, 111:166, 1996.
- [00284] Hua, L. (2000) Ph.D. Dissertation, Rice University, Houston, TX.
- [00285] Ishida *et al.*, "High efficiency transformation of maize (*Zea mays* L.) mediated by *Agrobacterium tumefaciens*," Nat Biotechnol., 14:745-50, 1996.
- [00286] Ito, H., Fukuda, Y., Murata, K., and Kimura, A. (1983) J. Bacteriol. 153, 163-168.
- [00287] Kaeppeler *et al.*, Plant Cell Reports 9: 415-418, 1990.
- [00288] Kaeppeler *et al.*, "Silicon carbide fiber-mediated stable transformation of plant cells," Theor. Appl. Genet., 84(5-6):560-566, 1992.
- [00289] Keegstra, K., and Olsen, L. J. (1989) Annu. Rev. Plant Physiol. Plant Mol. Biol. 40, 471-501.
- [00290] Klee *et al.*, Bio-Technology, 3(7):637-642, 1985.
- [00291] Knittel *et al.*, Plant Cell Reports, 14(2-3):81-86, 1994.
- [00292] Kutney, J. P., and Han, K. (1996) Rec. Trav. Chim. Pays-Bas 115, 77-93.
- [00293] Le Bars PL, Katz MM, Berman N, Itil TM, Freedman AM, Schatzberg AF. (1997) A placebo-controlled, double-blind, randomized trial of an extract of *Ginkgo biloba* for dementia. North American EGB Study Group. J. Amer. Med. Assoc. 278(16):1327-32.
- [00294] Lazzeri, "Stable transformation of barley via direct DNA uptake. Electroporation- and PEG- mediated protoplast transformation," Methods Mol. Biol., 49:95-106, 1995.
- [00295] Lee *et al.*, Korean J. Genet., 11(2):65-72, 1989.

- 10044007-010702
- [00296] Li, T.-S., Li, J.-T., and Li, H.-Z. (1995) *J. Chromatogr., A* 715, 372-375.
- [00297] Liu, H., and Krizek, J. (1992) *Genetics* 132, 665-673.
- [00298] Lorz *et al.*, *Mol Gen Genet*, 199:178-182, 1985.
- [00299] Maas *et al.*, "Preparation and transformation of monocot protoplasts," *Methods Cell Biol.*, 50:383-399, 1995.
- [00300] Marcotte *et al.*, *Nature*, 335:454, 1988.
- [00301] McCabe, Martinell, *Bio-Technology*, 11(5):596-598, 1993.
- [00302] McCormac *et al.*, *Euphytica*, v. 99 (1) p. 17-25 :. 1998.
- [00303] Nagatani *et al.*, "DNA delivery into rice cells and transformation using silicon carbide whiskers," *Biotech. Tech.*, 11(7):471-473, 1997.
- [00304] Nakanishi, K. *Pure Appl. Chem.* 7. 89 (1967).
- [00305] Nakanishi, K. *Tetrahedron Letter*, 4. 299 (1967).
- [00306] Neau, E., Cartayrade, A., Balz, J-P., Carde, J-P., and Walter, J. (1997) *Plant Physiol. Biochem.* 35, 869-879.
- [00307] Nielsen, H., Engelbrecht, J., Brunak, S., and von Heijne, G. (1997) *Protein Eng.* 10, 1-6.
- [00308] Omirulleh *et al.*, "Activity of a chimeric promoter with the doubled CaMV 35S enhancer element in protoplast-derived cells and transgenic plants in maize." *Plant Mol. Biol.*, 21:415-28, 1993.
- [00309] Okabe, K. *et al. J. Chem. Soc.*(5) 2201 (1967).
- [00310] Peters, R.J., Flory, J.E., Jetter, R., Ravn, M.M., Lee, H.-J., Coates, R.M., Croteau, R.B. (2000) Abietadiene synthase from Grand Fir (*Abies grandis*): characterization and mechanism of action of the "Pseudomature" recombinant enzyme. *Biochem.* 39: 15592-15602.
- [00311] Potrykus *et al.*, *Mol. Gen. Genet.*, 199:183-188, 1985.
- [00312] Ravn, M. M., Coates, R. M., Jetter, R., and Croteau, R. (1998) *Chem. Commun.* 21-22.
- [00313] Ravn, M. M., Coates, R. M., Flory, J. E., Peters, R. J., and Croteau, R. (2000) *Org. Letts.* 2, 573-576.



- [00314] Rhodes *et al.*, "Transformation of maize by electroporation of embryos," *Methods Mol. Biol.*, 55:121-131, 1995.
- [00315] Ritala *et al.*, "Fertile transgenic barley to particle bombardment of immature embryos," *Plant Mol. Biol.*, 24(2):317-325, 1994.
- [00316] Rogers *et al.*, *Methods Enzymol.*, 153:253-277, 1987.
- [00317] Ruan, B. (1999) Ph.D. Dissertation, Rice University, Houston, Texas, 1999.
- [00318] Sambrook, Fritsch, Maniatis, *In: Molecular Cloning: A Laboratory Manual*, Vol. 1, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, Ch. 7, 19-17.29, 1989.
- [00319] Schenk, P.M., Baumann, S., Mattes, R., Steinbiss, H. (1995) *BioTechniques* 19:196-200.
- [00320] Schwarz, M. (1994) Ph.D. Dissertation, E.T.H., Zurich, Switzerland, 1994.
- [00321] Schwarz, M., and Arigoni, D. (1999) *Comp. Nat. Prod. Chem.* 2, 367-400.
- [00322] Shimoni *et al.*, "A recombinant protein of two high molecular weight glutenins alters gluten polymer formation in transgenic wheat," *J. Biol. Chem.*, 272:15488-15495, 1997.
- [00323] Singsit *et al.*, "Expression of a *Bacillus thuringiensis* cryIA(c) gene in transgenic peanut plants and its efficacy against lesser cornstalk borer," *Transgenic Res.*, 6:169-76, 1997.
- [00324] Starks, C. M., Back, K., Chappell, J., and Noel, J. P. (1997) *Science* 277, 1815-1820.
- [00325] Takumi and Shimada, "Variation in transformation frequencies among six common wheat cultivars through particle bombardment of scutellar tissues," *Genes Genet. Sys.*, 72:63-69, 1997.
- [00326] Thompson *et al.*, "Maize transformation utilizing silicon carbide whiskers: A review," *Euphytica*, 85(1-3):75-80, 1995.
- [00327] Tingay *et al.*, *The Plant Journal* v. 11 (6) p. 1369-1376. 1997.

- [00328]       Tomes *et al.*, "Transgenic tobacco plants and their progeny derived by microprojectile bombardment of tobacco leaves," *Plant Mol. Biol.*, 14:261-8, 1990
- [00329]       Torbet *et al.*, "Transformation of oat using mature embryo-derived tissue cultures," *Crop Science*, 38:226-231, 1998.
- [00330]       Torbet *et al.*, "Use of paromomycin as a selective agent for oat transformation," *Plant Cell Reports*, 14:635-640, 1995.
- [00331]       Toriyama *et al.*, *Theor Appl. Genet.*, 73:16, 1986.
- [00332]       Tsukada; Kusano; Kitagawa, *Plant Cell Physiol.*, 30(4):599-604, 1989.
- [00333]       Uchimiya *et al.*, *Mol. Gen. Genet.*, 204:204, 1986.
- [00334]       Van Eck; Blowers; Earle, *Plant Cell Reports*, 14(5):299-304, 1995.
- [00335]       Vasil *et al.*, *Plant Physiol.*, 91:1575-1579, 1989.
- [00336]       von Heijne, G., and Nishikawa, K. (1991) *FEBS* 278, 1-3.
- [00337]       von Heijne, G., and Gavel, Y. (1990) *FEBS* 261, 455-458.
- [00338]       Walter, J. (1988) *Thèse d'Etat*, Université Bordeaux I, Bordeaux, France.
- [00339]       Williams, D. C., Wildung, M. R., Jin, A. Q., Dalal, D., Oliver, J. S., Coates, R. M., and Croteau, R. (2000) *Archives Biochem. Biophys.* 379, 137-146.
- [00340]       Yamada *et al.*, *Plant Cell Rep.*, 4:85, 1986.
- [00341]       Zhang *et al.*, "Agrobacterium-mediated transformation of elite indica and japonica rice cultivars," *Mol. Biotechnol.*, 8(3):223-231, 1997.
- [00342]       Zheng and Edwards, "Expression of resistance to barley stripe mosaic virus in barley and oat protoplasts," *J. Gen. Virol.*, 71:1865-1868, 1990.
- [00343]       Zhou *et al.*, *Plant Cell Reports*, 12(11):612-616, 1993.

[00344]       One skilled in the art readily appreciates that the patent invention is well adapted to carry out the objectives and obtain the ends and advantages mentioned as well as those inherent therein. Plant cells, yeast cells, cell cultures, plants, sequences, methods, procedures and techniques described herein are presently representative of the preferred embodiments and are intended to be exemplary and are not intended as limitations of the scope. Changes therein and other uses will occur to those skilled in the

art which are encompassed within the spirit of the invention or defined by the scope of the pending claims.

10041007.010702